

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

**Yu, Misook**

---

**From:** Mosher, Mary  
**Sent:** Tuesday, August 19, 2003 7:22 AM  
**To:** Yu, Misook  
**Subject:** RE: 09765086

Your judgement is better than mine in deciding whether or not there really is motivation to put the peices together, since you have read the references and I have not. If you, in your scientist role, sat down with the references, would it have occurred to you to put them together to get the invention, if you did not already know about the invention? If the invention is not a method of cancer treatment, is there an enablement issue on how to use the claimed invention for purposes other than treating cancer?

Sorry for the socratic response, but I really can't substitute my ignorance for your judgement.

MM

-----Original Message-----

**From:** Yu, Misook  
**Sent:** Monday, August 18, 2003 12:38 PM  
**To:** Mosher, Mary  
**Subject:** 09765086

There are 3 sets of method claims using a chimeric construct consisting of two parts (one, prostate-targeting peptide and the other, lysis-causing anti-microbial peptide).

The first set of claim says "a method of directing an antimicrobial peptide in vivo to prostate tissue";

The second set of claims says "a method of inducing apoptosis in prostate tissue in vivo"; and

The third set of claims says "a method of selectively inducing apoptosis in normal prostate tissue"

The first and second sets were rejected by combination of primary reference showing prostate-targeting peptides are known in the art to direct cancer treating agents to prostate, by two types of secondary reference showing that there is a need in the prostate cancer treatment art for a method to reduce generalized side effects, and antimicrobial peptide has been used for cancer treatment. Combination of the primary and secondary references teach all necessary components of the products for directing something to prostate or inducing apoptosis in prostate.

In reply to OA, applicant argues that their invention is not drawn to method of cancer treatment and there is no motivation why one of ordinary skill would be motivated to do the purpose specified in the preamble of the claims. The most of the specification is about prostate cancer treatment using the chimeric product. Do I drop the rejection?

Examiner Misook Yu, Ph.D.  
703-308-2454 (Phone)  
Art Unit 1642  
CM1-8E18 (Room)  
CM1-8E12 (Mail Box)





xx conditions -  
xx  
xx Claim 1; Page 151; 193pp; English.  
xx  
xx The present invention describes peptides that selectively home to a  
xx tissue or organ. The peptides can be used for identifying an organ  
xx or tissue, for identifying a target molecule expressed by an organ or  
xx tissue or for treating an organ or tissue pathology, where the organ or  
xx tissue is selected from prostate, lung, skin, retina, pancreas, gut,  
xx ovary, adrenal gland, liver, and lymph node. The peptide bind to the  
xx membrane dipeptidase (MDP). AA148618 to AA149066 represent sequences  
xx which are used in the exemplification of the present invention.

xx Sequence 7 A:

SQ

Query Match 100.0%; Score 30; DB 20; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

1 SMSIARL 7  
|||||||  
Db 1 smsiarl 7

RESULT 2  
AAB21907  
ID AAB21907 standard; Peptide; 7 AA.  
xx  
xx AAB21907;  
xx  
xx DT 22-MAR-2001 (first entry)  
xx  
xx DE Human prostate-homing peptide #1.  
xx  
xx CYC Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
xx KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; human.  
xx  
xx OS Homo sapiens.  
xx  
xx PN WO200042973-A2.  
xx  
xx PD 27-JUL-2000.  
xx  
xx PF 21-JAN-2000; 2000WO-US01602.  
xx  
xx PR 22-JAN-1999; 99US-0235902.  
xx  
xx RA (BURN-) BURNHAM INST.  
xx  
xx EL Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;  
xx WPI: 2000-499174/44.  
xx  
xx DR Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
xx PT selectively homes to a mammalian cell type or tissue linked to an  
xx PI antimicrobial peptide, useful for the treatment of prostate cancer -  
xx PS  
xx  
xx Claim 25; Page 107; 118pp; English.

xx The present invention relates to homing pro-apoptotic conjugates,  
xx comprising of a tumour homing molecule that selectively homes to a  
xx CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
xx CC homing pro-apoptotic conjugates are selectively internalised by the  
xx CC mammalian cell type or tissue and exhibits high toxicity, especially to  
xx CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
xx CC toxicity when not linked to the tumor homing molecule. The conjugates are  
xx CC useful for the treatment of cancer e.g. Kaposi's sarcoma, breast and  
xx CC prostate cancer or melanoma. The present sequence is a homing peptide  
xx CC isolated in the present invention, which can be conjugated to an  
xx CC antimicrobial peptide to make the homing pro-apoptotic conjugates of the  
xx CC present invention.

SQ	Sequence	7 AA;
OY	1 SMSIARL 7       1 smsiarl 7	Query Match Best Local Similarity 100.0%; Score 30; DB 21; Length 7; Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;
RESULT	3	
ID	AAE06483	
AC	AAE06483 standard; peptide; 7 AA.	
AE	AAE06483;	
DT	25-SEP-2001 (first entry)	
DE	Mouse prostate homing peptide #1.	
KW	Mouse; chimeric prostate-homing pro-apoptotic peptide;	
RK	prostate-homing peptide; antimicrobial peptide; prostate cancer;	
KM	tumour homing molecule; cytostatic.	
OS	Mus sp.	
PX	WO20015342-A1.	
PN	26-JUL-2001.	
PF	16-JAN-2001; 2001WO-US01362.	
PR	21-JAN-2000; 2000US-0489582.	
PA	(BURN-) BURNHAM INST.	
PI	Ruoslahti EI, Pasqualini R, Arap W, Bredesen DE, Ellnerby HM;	
DR	WPI; 2001-451901/48.	
PT	Novel chimeric prostate-homing pro-apoptotic peptide, used to treat	
PT	prostate cancer, comprises a prostate-homing peptide linked to an	
PS	antimicrobial peptide -	
XX	Claim 2; Page 103; 176pp; English.	
CC	The patent discloses novel chimeric prostate-homing pro-apoptotic	
CC	peptide which comprises a prostate-homing peptide linked to an	
CC	antimicrobial peptide, where the chimeric peptide is selectively	
CC	internalised by and exhibits high toxicity to prostate tissue and	
CC	where the antimicrobial peptide has low mammalian cell toxicity when	
CC	not linked to prostate-homing peptide. The chimeric peptide is used	
CC	to direct an antimicrobial peptide in vivo to a prostate cancer, to	
CC	induce selective toxicity in vivo in a prostate cancer, and to treat	
CC	a patient with prostate cancer. The present sequence is mouse prostate	
CC	homing peptide. This sequence is useful in the homing of pro-apoptotic	
CC	conjugates of the invention.	
SO	Sequence 7 AA;	
OY	1 SMSIARL 7       1 smsiarl 7	Query Match Best Local Similarity 100.0%; Score 30; DB 22; Length 7; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	1 smsiarl 7	
RESULT	4	

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:18:34 ; Search time 38.66 seconds

(without alignments)  
13,412 Million cell updates/sec

Title: US-09-765-086-207

Perfect score: 30

Sequence: 1 SMSIARL 7

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

1: /SID58/gcgdata/geneseq/AA1980.DAT:\*  
2: /SID58/gcgdata/geneseq/AA1981.DAT:\*  
3: /SID58/gcgdata/geneseq/AA1982.DAT:\*  
4: /SID58/gcgdata/geneseq/AA1983.DAT:\*  
5: /SID58/gcgdata/geneseq/AA1984.DAT:\*  
6: /SID58/gcgdata/geneseq/AA1985.DAT:\*  
7: /SID58/gcgdata/geneseq/AA1986.DAT:\*  
8: /SID58/gcgdata/geneseq/AA1987.DAT:\*  
9: /SID58/gcgdata/geneseq/AA1988.DAT:\*  
10: /SID58/gcgdata/geneseq/AA1989.DAT:\*  
11: /SID58/gcgdata/geneseq/AA1990.DAT:\*  
12: /SID58/gcgdata/geneseq/AA1991.DAT:\*  
13: /SID58/gcgdata/geneseq/AA1992.DAT:\*  
14: /SID58/gcgdata/geneseq/AA1993.DAT:\*  
15: /SID58/gcgdata/geneseq/AA1994.DAT:\*  
16: /SID58/gcgdata/geneseq/AA1995.DAT:\*  
17: /SID58/gcgdata/geneseq/AA1996.DAT:\*  
18: /SID58/gcgdata/geneseq/AA1997.DAT:\*  
19: /SID58/gcgdata/geneseq/AA1998.DAT:\*  
20: /SID58/gcgdata/geneseq/AA1999.DAT:\*  
21: /SID58/gcgdata/geneseq/AA2000.DAT:\*  
22: /SID58/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	7	AAV48905	Membrane dipeptidase
2	30	100.0	7	AAAB21907	Human prostate hom
3	30	100.0	7	AAE06483	Mouse prostate hom
4	30	100.0	23	AAAB21938	Prostate homing an
5	30	100.0	23	AAE06513	Chimeric prostate-
6	29	96.7	23	AAAG90583	C glutamicum prote
7	27	90.0	21	AAAG08955	Arabidopsis thalia
8	27	90.0	289	AAAG91999	C glutamicum prote
9	27	90.0	301	AAAG34469	Protein involved i
10	26	86.7	183	AAU14179	Human novel protei
11	26	86.7	260	AAAB1860	Human ORFX ORF1624

12	26	86.7	382	22	AAAG92195	C glutamicum prote
13	26	86.7	382	22	AAAB9265	Corynebacterium g1
14	26	86.7	422	18	AAW18032	zebrafish retinoid
15	26	86.7	697	21	AAAG25565	Eucalyptus grandis
16	26	86.7	748	22	AAW40225	Human polypeptide
17	25	83.3	118	22	AAAG8415	Escherichia coli p
18	25	83.3	118	22	AAAG8916	E. coli growth and
19	25	83.3	118	22	AAAG8948	E. coli growth and
20	25	83.3	204	21	AAV75241	Neisseria gonorrhe
21	25	83.3	381	22	AAAG92942	C glutamicum prote
22	24	80.0	64	21	AAAB9477	Human secreted pro
23	24	80.0	120	22	AAAB94277	Human secreted pro
24	24	80.0	137	18	AAW20100	Human protein sequ
25	24	80.0	266	19	AAW88839	H. pylori cytoplas
26	24	80.0	319	21	AAAG04867	H. pylori GPO 153
27	24	80.0	319	21	AAAG42777	Arabidopsis thalia
28	24	80.0	336	21	AAAG04866	Arabidopsis thalia
29	24	80.0	336	21	AAAG42776	Arabidopsis thalia
30	24	80.0	342	21	AAAG04865	Arabidopsis thalia
31	24	80.0	363	21	AAAG42775	Arabidopsis thalia
32	24	80.0	363	21	AAAG21936	Arabidopsis thalia
33	24	80.0	381	21	AAAG21935	Arabidopsis thalia
34	24	80.0	387	21	AAAG21934	Arabidopsis thalia
35	24	80.0	477	15	AAAG62448	Arabidopsis thalia
36	24	80.0	515	15	AAAG62447	Rat calcitonin rec
37	24	80.0	2675	21	AAAB07564	Rat calcitonin rec
38	23	76.7	23	21	AAAB27643	Protein encoded by
39	23	76.7	42	11	AAAR07414	Human secreted pro
40	23	76.7	69	21	AAAB44318	PJD5-neo-ATP trans
41	23	76.7	63	21	AAAG55635	Human secreted pro
42	23	76.7	63	21	AAAG58828	Human secreted pro
43	23	76.7	81	22	AAAM14873	Arabidopsis thalia
44	23	76.7	81	22	AAAM27300	Peptide #1307 enco
45	23	76.7	81	22	AAAM02598	Peptide #1280 enco

#### ALIGNMENTS

RESULT 1	
AAV48905	AAV48905 standard; Peptide: 7 AA.
AC	AAV48905;
XX	
XX	10-DEC-1999 (first entry)
DE	Membrane dipeptidase-binding prostate homing peptide #21.
XX	
XX	Homing peptide; organ; tissue; lung; pancreas; skin; retina; MDP;
KW	prostate; ovary; lymph node; adrenal gland; liver; gut; tumour;
KM	membrane dipeptidase.
XX	
OS	Synthetic.
OS	Homoe-septens
XX	
XX	MO9946284-A2
XX	16-SEP-1999
XX	
PF	10-MAR-1999; 99WO-0505284.
XX	
PR	13-MAR-1998; 98US-0042107.
PR	26-FEB-1999; 99US-0042107.
PA	(BURN-) BURNHAM INST.
XX	
PI	Rajotte D, Pasqualini R, Ruoslahti EI;
XX	
DR	WPI: 1999-571717/48.
XX	
XX	New peptides which selectively home to organs or tissues, used for,
PT	e.g. identifying target ligands and for therapy of pathological
PT	

Best Local Similarity 100.0%; Pred. No. 0.0029;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 K1AK1AK1AK1AK 14  
Db 1 K1AK1AK1AK1AK 14

RESULT 2  
US-08-944-133-3  
Sequence 3, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L  
TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Runnels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Runnels, John H  
REGISTRATION/DOCKET NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELEPHONE: 504 346-8049  
TELEFAX: 504 387-3221  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-3

Query Match 100.0%; Score 62; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 K1AK1AK1AK1AK 14  
Db 5 K1AK1AK1AK1AK 18

RESULT 3  
US-08-944-133-7  
Sequence 7, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L

TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Runnels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Runnels, John H  
REGISTRATION/DOCKET NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELEPHONE: 504 346-8049  
TELEFAX: 504 387-3221  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-7

Query Match 100.0%; Score 62; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 K1AK1AK1AK1AK 14  
Db 1 K1AK1AK1AK1AK 14

RESULT 4  
US-08-944-133-26  
Sequence 26, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L  
TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Runnels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:17:52 ; Search time 20.74 Seconds  
(without alignments)  
15.190 Million cell updates/sec

Title: US-09-765-086-200

Perfect score: 62

Sequence: 1 KLAKLAKLAKLAK 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252.seqs, 22503292 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/prodata/2/1aa/PCFUS.COMB.pep:\*
- 6: /cgn2\_6/prodata/2/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	62	100.0	14 1 US-08-944-133-6	Sequence 6, Appl
2	62	100.0	21 1 US-08-944-133-3	Sequence 3, Appl
3	62	100.0	21 1 US-08-944-133-7	Sequence 7, Appl
4	62	100.0	21 1 US-08-944-133-26	Sequence 26, Appl
5	62	100.0	28 1 US-08-944-133-4	Sequence 4, Appl
6	62	100.0	28 1 US-08-944-133-8	Sequence 8, Appl
7	62	100.0	28 1 US-08-944-133-27	Sequence 27, Appl
8	62	100.0	28 1 US-08-944-133-25	Sequence 25, Appl
9	57	91.9	23 2 US-08-723-306-27	Sequence 27, Appl
10	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
11	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
12	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
13	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
14	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
15	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
16	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
17	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
18	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
19	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
20	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
21	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
22	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
23	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
24	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
25	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
26	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl
27	57	91.9	23 2 US-08-723-306-28	Sequence 28, Appl

## ALIGNMENTS

28	52	83.9	31 5	PCT-US94-12550-48	Sequence 48, Appl
29	52	83.9	31 5	PCT-US96-10041-26	Sequence 26, Appl
30	52	83.9	38 2	US-08-723-306-23	Sequence 23, Appl
31	52	83.9	38 2	PCT-US96-10041-23	Sequence 23, Appl
32	50	80.6	21 1	US-07-908-455A-4	Sequence 4, Appl
33	50	80.6	21 1	US-07-908-455A-64	Sequence 64, Appl
34	50	80.6	21 1	US-08-434-120-30	Sequence 30, Appl
35	50	80.6	21 1	US-08-434-120-30	Sequence 30, Appl
36	50	80.6	21 1	US-08-465-325-30	Sequence 30, Appl
37	50	80.6	21 1	US-08-465-325-88	Sequence 88, Appl
38	50	80.6	23 1	US-08-231-730A-24	Sequence 24, Appl
39	50	80.6	23 1	US-08-427-001C-24	Sequence 24, Appl
40	50	80.6	23 1	US-08-457-798-24	Sequence 24, Appl
41	50	80.6	23 1	US-08-457-171-24	Sequence 24, Appl
42	50	80.6	23 2	US-08-505-466-24	Sequence 24, Appl
43	50	80.6	23 3	US-08-689-489C-24	Sequence 24, Appl
44	50	80.6	23 3	US-08-801-028-24	Sequence 24, Appl
45	50	80.6	23 5	PCT-US94-06176-24	Sequence 24, Appl

RESULT 1  
US-08-944-133-6  
Sequence 6, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L  
TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESS: John H. Runnels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Runnels, John H  
REGISTRATION NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 504 387-3221  
TELEFAX: 504 346-8049  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-6  
Query Match 100.0%; Score 62; DB 1; Length 14;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 13:34:08 ; Search time 10.17 Seconds  
(without alignments)  
82.919 Million cell updates/sec

Title: CHIMERA  
Perfect score: 104  
Sequence: 1 SMSIARLGKRLAKLAKLAK 23

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	47	45.2	1	Q9ZG88 Caulobacter
2	45	43.2	362	P32779 BORHE
3	45	43.3	462	Y445 CHV1
4	43	41.3	756	FILS BOVIN
5	43	41.3	790	TOPI_DAUCA
6	43	41.3	1722	RBB2_HUMAN
7	42	40.4	273	RK2_MAIZE
8	42	40.4	293	VHED_BPT2
9	42	40.4	293	VHED_BPT6
10	42	40.4	352	NPTI_ASPOR
11	42	40.4	641	DNK_METSS
12	42	40.4	1188	TIRK_ECOLI
13	41.5	39.9	548	PPOC_TOBAC
14	41	39.4	112	HMGD_DROME
15	41	39.4	216	Y458 METJA
16	41	39.4	221	YM14_ARCFU
17	41	39.4	242	OMP4_NEITMA
18	41	39.4	254	YCAL_ECOLI
19	41	39.4	364	YH39_BORHE
20	41	39.4	368	YH39_ARCFU
21	41	39.4	431	SAHH_DROME
22	41	39.4	460	EF1A_YARLI
23	41	39.4	485	SAHH_MESCR
24	41	39.4	485	SAHH_PETCR
25	41	39.4	485	SAHH_TOBAC
26	41	39.4	645	SHTE_DROME
27	41	39.4	945	SYV_NEITMA
28	40.5	38.9	716	P5C5_ORYSA
29	40	38.5	116	RL19_VIBCH
30	40	38.5	168	RL17_STRCO
31	40	38.5	248	DJ_DROME
32	40	38.5	301	VHED_BPT4
33	38.5	319	1	YHAI_CRYPA

34	40	38.5	413	1	PROA_LACIA	Q9CF73 Lactococcus
35	40	38.5	416	1	Y748_METJA	O58158 methanococ
36	40	38.5	437	1	SAHH_LEIDO	P36889 leishmania
37	40	38.5	448	1	Y260_STRY3	F74409 synectocyst
38	40	38.5	453	1	TRME_BOCMP	O51830 buchnera ap
39	40	38.5	459	1	EF1A_CRYNE	O42671 cryptococcu
40	40	38.5	485	1	SAHH_CATRO	P35007 catharanthu
41	40	38.5	485	1	SAHH_MDSA	P50246 medicago sa
42	40	38.5	520	1	YMDA_BACSU	O31774 bacillus su
43	40	38.5	539	1	DOP2_DROME	O24563 dirosophila
44	40	38.5	560	1	ARAB_BACSU	P94524 bacillus su
45	40	38.5	909	1	CNG4_HUMAN	Q14028 homo sapien

## ALIGNMENTS

RESULT ID	1	PODI_CAUCR	STANDARD;	PRT;	472 AA.
AC	Q9ZG88;	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	POLAR ORGANELLE DEVELOPMENT PROTEIN.				
GN	PODI OR CC2045.				
OS	Caulobacter crescentus.				
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;				
OC	Caulobacter.				
OX	NCBI_TaxID=69394;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 19089 / CB15;				
RX	MEDLINE=21173698; PubMed=10383964;				
RA	Crymes W.B. Jr., Zhang D., Ely B.;				
RT	"Regulation of podi expression during the Caulobacter crescentus cell cycle.";				
RL	J. Bacteriol. 181:3967-3973(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 19089 / CB15;				
RX	MEDLINE=21173698; PubMed=11259647;				
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.K.R., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.R., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Shapiro L., Venter J.C., Fraser C.M.;				
RT	"Complete genome sequence of Caulobacter crescentus.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).				
CC	-!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
CC	EMBL: AF084609; AAC72820.1; -;				
DR	EMBL; AE005877; -; NOT_ANNOTATED_CDS.				
DR	TIGR; CC2045; -;				
KW	Complete proteome.				
FT	DOMAIN 432				
FT	CONFICT 254 255				
FT	CONFICT 428 428				
FT	CONFICT 457 457				
SQ	SEQUENCE 472 AA; 52422 MW; 3F064FB9B4E413D CRC64;				

```

Query Match 45.2%; Score 47; DB 1; Length 472;
Best Local Similarity 35.0%; Pred. No. 15;
Matches 7; Conservative 10; Mismatches 3; Indels 0; Gaps 0;

QY 1 SMSIARLGKTLAKTLAKLAK 20
      ::::|||||:::|||||
DB 339 AQALEKLGGEIARITEKRLAE 358

RESULT 2
ID VM25_BORHE STANDARD; PRT; 362 AA.
AC P33779;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE VARIABLE MAJOR OUTER MEMBRANE LIPOPROTEIN 25 PRECURSOR.
GN VME25.
OS Borrelia hermsli.
OC Plasmid.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_Taxid=140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. HSI SEROTYPE 25;
RA MEDLINE=931133110; Pubmed=1484486;
RA Restrepo B.I., Kitten T., Carter C.J., Infante D., Barbours A.G.;
RT "Subtelomeric expression regions of Borrelia hermsli linear plasmids
RT are highly polymorphic.";
RL Mol. Microbiol. 6:329-331(1992).
CC -1- FUNCTION: SERVES TO AVOID THE HOST IMMUNE RESPONSE BY CHANGING
CC FROM ONE SURFACE EXPOSED VMP TO ANOTHER.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR.
CC -1- SIMILARITY: STRONG, TO VMEP1 AND VMP21.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L04787; AAA22965.1; -
CC DR InterPro; IPR000680; Borrelia_lipo.
CC DR Pfam; PF00921; Lipoprotein_2; 1.
CC DR PROSITE; PS00013; PROKAR LIPOPROTEIN. 1.
CC KW Outer membrane; Lipoprotein; Signal; Plasmid.
CC FT SIGNAL 1 26 PROBABLE.
CC FT CHAIN 27 362 VARIABLE MAJOR OUTER MEMBRANE
CC FT LIPID 27 27 N-ACYL DIGLYCERIDE (PROBABLE).
CC FT SEQUENCE 362 AA; 36468 MW; 2A3EF5248D5637FD CRC64;

Query Match 43.3%; Score 45; DB 1; Length 362;
Best Local Similarity 45.5%; Pred. No. 22;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 SMSIARLGKTLAKTLAKTLAKLA 22
      :::|1111::|111::111
DB 268 ALKFARGGNGQLAKKEAKAA 289

RESULT 3
ID Y445_CHVP1 STANDARD; PRT; 462 AA.
AC Q98496;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN A445L.

```

```

OS A445L.
OS Paramesicium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX NCBI_TaxID=10506;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=96400190; PubMed=8806566;
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map
RL positions 182 to 258."
RL Virology 223:303-317(1996).
CC -1- SIMILARITY: BELONGS TO THE ABC1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U42580; AAC96813.1; -.
KW Hypothetical protein.
SQ SEQUENCE 462 AA; 52541 MW; 77BFC15A7EF9EDC7 CRC64;

Query Match 43.3%; Score 45; DB 1; Length 462;
Best Local Similarity 36.4%; Pred. No. 27;
Matches 8; Conservative %; 6; Mismatches 8; Indels 0; Gaps 0;

OY 2 MSIALGSKLAKLAKLAKLAK 23
|::|::|::|::|::|
Db 309 MGVIKSGSVARLKAIPKIRK 330

RESULT 4
FILS_BOVIN STANDARD; PRT; 756 AA.
AC Q06002;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FILENSIN (BEADED FILAMENT STRUCTURAL PROTEIN 1).
GN BEP1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eumetazoa; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Lens;
RC MEDLINE=93260017; PubMed=8491777;
RA Gounaris C.A., Merdes A., Quinlan R., Hess J.F., Fitzgerald P.G.,
RA Gounaris C.A., Georgatos S.D.;
RT "Bovine filensin possesses primary and secondary structure similarity
RT to intermediate filament proteins."
RL J. Cell Biol. 121:847-853(1993).
RN [2]
RN REVISIONS, AND SEQUENCE FROM N.A.
RP TISSUE=Lens;
RC Hess J.F.;
RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: MEMBRANE- AND CYTOSKELETON-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: LENS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: X72388; CAA51081.1; -.
CC DR InterPro: IPR001664; IF.
CC DR PROSITE: PS00226; IF: FALSE_NEG.
CC KW Intermediate filament; Repeat; Membrane; Coiled coil; Cytoskeleton;
CC Eye lens protein; Phosphorylation.
CC FT DOMAIN 1 38 HEAD.
CC FT DOMAIN 39 318 ROD.
CC FT DOMAIN 319 755 TAIL.
CC FT DOMAIN 39 73 COIL 1A.
CC FT DOMAIN 74 82 COIL 1B.
CC FT DOMAIN 83 182 LINKER 1.
CC FT DOMAIN 183 199 LINKER 12.
CC FT DOMAIN 200 318 COIL 2.
CC FT DOMAIN 531 621 7 x 14 AA TANDEM REPEATS.
CC FT REPEAT 531 544 1.
CC FT REPEAT 545 551 2 (INCOMPLETE).
CC FT REPEAT 552 565 3.
CC FT REPEAT 566 579 4.
CC FT REPEAT 580 593 5.
CC FT REPEAT 594 607 6.
CC FT REPEAT 608 621 7.
CC MOD_RES 5 5 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
CC SQ SEQUENCE 756 AA; 83001 MW; F86A18208A8E6109 CRC64;

Query Match 41.3%; Score 43; DB 1; Length 756;
Best Local Similarity 50.0%; Pred. No. 82;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 8 GGKIAKIAKIAKIAK 23
DB 432 GGIKIAKIAKIAK 447

RESULT 5
TOP1 DAUCA STANDARD; PRT; 790 AA.
AC P93119;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA TOPOISOMERASE I (EC 5.99.1.2).
GN TOP1.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LUNGA DI AMSTERDAM;
RX MEDLINE=97149297; PubMed=8996105;
RA Balistreri A., Toscano I., Bernacchia G., Luo M., Otte S.,
RA Caronera D.;
RT "Cloning of a cDNA encoding DNA topoisomerase I in Daucus carota and
RT expression analysis in relation to cell proliferation.";
RL Gene 183:183-190(1996).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U60440; AAB41401.1; -.
CC DR HSSP: P11387; 1A35.
CC DR Mendel; 17369; Daucus; 2787; 17369.
CC DR InterPro: IPR001631; Topoisomerase_I.
CC DR Pfam; PF01028; Topoisomerase_I; 1.
CC DR PRINTS: PRO0416; EGTPISMRASEI.
CC DR SMART: SM00435; TOPEUC; 1.
CC DR PROSITE: PS00176; TOPOISOMERASE_I_EUK; 1.
CC KW Isomerase; Topoisomerase; DNA-binding.
CC FT ACT_SITE 749 749 DNA CLEAVAGE (BY SIMILARITY).
CC SQ SEQUENCE 790 AA; 89770 MW; CE1D7DA3FF3291P5 CRC64;

Query Match 41.3%; Score 43; DB 1; Length 790;
Best Local Similarity 45.0%; Pred. No. 86;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 SMSIARLGKIAKIAK 20
DB 685 NLNKRKYKILKILK 704

RESULT 6
RBB2 HUMAN STANDARD; PRT; 1722 AA.
AC P29375;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RETINOBLASTOMA BINDING PROTEIN 2 (RBBP-2).
GN RBBP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94020841; PubMed=8414517;
RA Falcay A.R., Helin K., Dembski M.S., Dyson N., Harlow E.,
RA Vuocolo G.A., Hanobik M.G., Haskell K.M., Olf A., Defeo-Jones D.,
RA Jones R.E.;
RT "Characterization of the retinoblastoma binding proteins RBP1 and
RT RBP2.";
RL Oncogene 8:3149-3156(1993).
RN [2]
RP SEQUENCE OF 1102-1569 FROM N.A.
RX MEDLINE=91312450; PubMed=1857421;
RA Defeo-Jones D., Huang P.S., Jones R.E., Haskell K.M., Vuocolo G.A.,
RA Hanobik M.G., Huber H.E., Olf A.;
RT "Cloning of cDNAs for cellular proteins that bind to the
RT retinoblastoma gene product.";
RL Nature 352:251-254(1991).
CC -1- FUNCTION: INTERACTS WITH THE VIRAL PROTEIN-BINDING DOMAIN OF THE
CC RETINOBLASTOMA PROTEIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO THE C-TERMINAL DOMAIN OF THE XE169 PROTEIN AND
CC THE JUMONJI PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----

```

DR EMBL: S66431; AAB28544.1; -  
DR PIR: S16954; S16954.  
DR MIM: 180202; -  
DR InterPro: IPR001606; ARID.  
DR InterPro: IPR003347; JmjC.  
DR InterPro: IPR003349; JmjN.  
DR InterPro: IPR001965; PHD.  
DR Pfam: PF01388; ARID; 1.  
DR Pfam: PF02373; JmjC; 1.  
DR Pfam: PF02375; JmjN; 1.  
DR Pfam: PF00628; PHD; 3.  
DR SMART: SM00501; BRIGHT; 1.  
DR SMART: SM00545; JmjN; 1.  
DR SMART: SM00249; PHD; 3.  
KW Trans-acting factor; Nuclear protein.  
FT CONFLICT 1563 1566 MISSING (IN REF. 2).  
SQ SEQUENCE 1722 AA; 195815 MW; 8CF8A8A8E9A652 CRC64;

Query Match 41.3%; Score 43; DB 1; Length 1722;  
Best Local Similarity 90.9%; Pred. No. 1.8e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 10 KLAKLAKLAK 20  
DB 1552 KLAKLAKLAK 1562

RESULT 7  
RK2\_MAIZE STANDARD; PRT: 273 AA.  
ID RK2\_MAIZE  
AC P17788;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L2.  
GN RPL2.  
OS Zea mays (Maize).  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Palicoudeae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV\_F99CMS837;  
RX MEDLINE=90332419; PubMed=2377464;  
RA Kavousi M., Glese K., Lartina I.M., Subramanian A.R.;  
RT "Nucleotide sequence and map positions of the duplicated gene for  
maize (Zea mays) chloroplast ribosomal protein L2.";  
RL Nucleic Acids Res. 18:4244-4244(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95395841; PubMed=766415;  
RA Maier R.M., Neckermann K., Igloi G.L., Koessel H.;  
RT "Complete sequence of the maize chloroplast genome: gene content,  
hotspots of divergence and fine tuning of genetic information by  
transcript editing";  
RL J. Mol. Biol. 251:614-628(1995).  
RN [3]  
RP RNA EDITING OF INITIATOR CODON.  
RX MEDLINE=91367263; PubMed=1653905;  
RA Hoch B., Maier R.M., Appel K., Igloi G.L., Koessel H.;  
RT "Editing of a chloroplast mRNA by creation of an initiation codon";  
RL Nature 353:178-180(1991).  
CC -! SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.  
CC -! CAUTION: THE INITIATOR METHIONINE IS CREATED BY RNA EDITING.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X53066; CAA37241.1; -  
DR EMBL: X86563; CAA60329.1; ALT\_SEQ.  
DR EMBL: X86563; CAA60371.1; -  
DR EMBL: X62070; CAA43983.1; ALT\_SEQ.  
DR PIR: S10500; R52M2.  
DR PIR: S17874; S17874.  
DR MaizEDB: 66413; -  
DR Mendel: 11440; ZEMa:rp12.1.  
DR InterPro: IPR002171; Ribosomal\_L2.  
DR Pfam: PF00181; Ribosomal\_L2; 1.  
DR PROSITE: PS00467; RIBOSOMAL\_L2; 1.  
KW Ribosomal protein; Chloroplast; RNA editing.  
SQ SEQUENCE 273 AA; 30065 MW; BA65197231EA3CA0 CRC64;

Query Match 40.4%; Score 42; DB 1; Length 273;  
Best Local Similarity 52.6%; Pred. No. 45;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 4 IARIGKTLAKLAKLAKLA 22  
DB 152 IARAGAAVAKLAKKEGKLA 170

RESULT 8  
VHED\_BPT2 STANDARD; PRT: 293 AA.  
ID VHED\_BPT2  
AC P09035;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SINGLE-STRANDED DNA BINDING PROTEIN (HELIX-DESTABILIZING PROTEIN)  
DE (GP32) (FRAGMENT).  
GN 32 OR SSB.  
OS Bacteriophage T2.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;  
OC T4-like phages.  
OX NCBI\_TaxID=10664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89016637; PubMed=3262868;  
RA McPheters D.S., Stormo G.D., Gosh G., Gold L.;  
RT "Nucleotide sequences of the bacteriophage T2 and T6 gene 32 mRNAs";  
RL Nucleic Acids Res. 16:9341-9342(1988).  
CC -! FUNCTION: BINDS PREFERENTIALLY TO SINGLE-STRANDED DNA AND  
CC THEREFORE, DESTABILIZES DOUBLE-STRANDED DNA. IT IS INVOLVED IN DNA  
CC REPLICATION, REPAIR AND RECOMBINATION. BINDS SS-DNA AS THE  
CC REPLICATION FORK ADVANCES AND STIMULATES THE REPLICASOME  
CC PROCESSIVITY AND ACCURACY.  
CC -! SUBUNIT: HOMODIMER IN THE ABSENCE OF DNA, MONOMER WHEN BINDING  
CC DNA (BY SIMILARITY).  
CC -! MISCELLANEOUS: INTERACTS WITH THE POLYMERASE AND THE UVSX AND UVSX  
CC PROTEINS (BY SIMILARITY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X12460; CAA31000.1; -  
DR EMBL: X12460; CAA31001.1; ALT\_SEQ.  
DR PIR: S01437; DDBP32.  
DR HSSP: P03695; IGPC.  
KW Zinc-binding; DNA repair; DNA replication; DNA recombination;  
KW Zinc-finger; Zinc.  
FT METAL 65 65 ZINC (BY SIMILARITY).  
FT METAL 78 78 ZINC (BY SIMILARITY).



FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT METAL 91 91 ZINC (BY SIMILARITY).  
 FT NON\_TER 293 293  
 SQ SEQUENCE 293 AA; 32451 MW; F46AB095244D90B7 CRC64;

Query Match 40.4%; Score 42; DB 1; Length 293;  
 Best Local Similarity 47.6%; Pred. No. 48;  
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 2 MSTARLGKLAIAKLA 22  
 DB 240 MGTAAAGAAATRAKKADKVA 260

RESULT 9  
 VHEHD\_BPT6 STANDARD; PRT; 293 AA.  
 AC P09797;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SINGLE-STRANDED DNA BINDING PROTEIN (HELIX-DESTABILIZING PROTEIN)  
 DE (GP32). (FRAGMENT).  
 GN 32 OR SSB.  
 OS Bacteriophage T6.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;  
 OC T4-like phages.  
 OX NCBI\_Taxid=10666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89016637; PubMed=3262868;  
 RA McPheeters D.S., Stormo G.D., Gosh G., Gold L.;  
 RL Nucleotide sequences of the bacteriophage T2 and T6 gene 32 mRNAs.;  
 CC Nucleic Acids Res. 16:9341-9342(1988).  
 CC -1- FUNCTION: BINDS PREFERENTIALLY TO SINGLE-STRANDED DNA AND  
 THEREFORE, DESTABILIZES DOUBLE-STRANDED DNA. IT IS INVOLVED IN DNA  
 REPLICATION, REPAIR AND RECOMBINATION. BINDS SS-DNA AS THE  
 REPLICATION FORK ADVANCES AND STIMULATES THE REPLICOME  
 PROCESSIVITY AND ACCURACY.  
 CC -1- SUBUNIT: HOMODIMER IN THE ABSENCE OF DNA. MONOMER WHEN BINDING  
 DNA (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: INTERACTS WITH THE POLYMERASE AND THE UVSX AND UVSX  
 PROTEINS (BY SIMILARITY).  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL: X12488; CAA31011.1; -.  
 DR PIR: S02263; DDBP36.  
 DR HSSP: P03695; IGPC.  
 KW DNA-binding; DNA repair; DNA replication; DNA recombination;  
 KM Zinc-finger; Zinc.  
 FT METAL 65 65 ZINC (BY SIMILARITY).  
 FT METAL 78 78 ZINC (BY SIMILARITY).  
 FT METAL 88 88 ZINC (BY SIMILARITY).  
 FT METAL 91 91 ZINC (BY SIMILARITY).  
 FT NON\_TER 293 293  
 SQ SEQUENCE 293 AA; 32468 MW; 09EDDCCB82AF4F6 CRC64;

Query Match 40.4%; Score 42; DB 1; Length 293;  
 Best Local Similarity 47.6%; Pred. No. 48;  
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
 OY 2 MSTARLGKLAIAKLA 22  
 DB 240 MGTAAAGAAATRAKKADKVA 260

RESULT 10  
 NP11\_ASPOR STANDARD; PRT; 352 AA.  
 AC P46076;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE NEUTRAL PROTEASE II PRECURSOR (PC 3.4.24.39) (DEUTEROLYSIN) (NP11).  
 OS Aspergillus oryzae.  
 OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_Taxid=5062;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 176-210; 279-281 AND 304-341.  
 RC STRAIN=ATCC 20386;  
 RX MEDLINE=91360097; PubMed=1886621;  
 RA Tatsumi H., Murakami S., Tsuji R.F., Ishida Y., Murakami K.,  
 RA Masaki A., Kawabe H., Arimura H., Nakano E., Motai H.;  
 RL "Cloning and expression in yeast of a cDNA clone encoding Aspergillus  
 oryzae neutral protease II, a unique metalloprotease.";  
 RL Mol. Gen. Genet. 228:97-103(1991).  
 CC -1- FUNCTION: THERMOSTABLE METALLOPROTEASE. SHOWS HIGH ACTIVITIES ON  
 BASIC NUCLEAR SUBSTRATES SUCH AS HISTONE AND PROTAMINE.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF BONDS WITH  
 HYDROPHOBIC RESIDUES IN P1', ALSO 3-ASN-1-GLN-4 AND 8-GLN-1-SRR-9  
 BONDS IN INSULIN B CHAIN.  
 CC -1- COFACTOR: BINDS ONE ZINC ION.  
 CC -1- PTM: PROBABLY POSSESSES THREE DISULFIDE BONDS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M35 (ZINC  
 METALLOPROTEASE); ALSO KNOWN AS THE DEUTEROLYSIN SUBFAMILY.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL: S53810; AAB19701.1; -.  
 DR MEROPS: M35.002; -.  
 DR InterPro: IPR001384; Peptidase\_M35.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF02102; Peptidase\_M35; 1.  
 DR PRINTS: PR00768; DEUTEROLYSIN.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Zymogen.  
 FT STGNAL 1 19 POTENTIAL.  
 FT PROPRP 20 175  
 FT CHAIN 176 352  
 FT METAL 303 303 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 304 304 BY SIMILARITY.  
 FT METAL 307 307 ZINC (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 352 AA; 37517 MW; 070C5131335B7F44 CRC64;

Query Match 40.4%; Score 42; DB 1; Length 352;  
 Best Local Similarity 50.0%; Pred. No. 57;  
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 OY 8 GGRKLAIAKLAIAK 23  
 DB 159 GGRKAVYKALISOLTR 174

RESULT 11  
 DNK1\_METSS STANDARD; PRT; 641 AA.  
 AC G92FC6;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CHAPERONE PROTEIN DNAK (HEAT SHOCK PROTEIN 70 KDA  
DE PROTEIN) (HSP70).  
CN DNAK.  
OS Methylovorus sp. (strain SSI / DSM 11726).  
OC Bacteria; Proteobacteria; beta subdivision; Methylophilus group;  
OC Methylovorus.  
OX NCBI\_TaxId=81683;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Eom C.Y., Kim Y.M.,  
RT "gprc, dnaK, and dnaJ genes of Methylovorus sp. strain SSI DSM11726.";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/1sb-sib.ch>).  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC -----  
DR EMBL; AF106835; AAC95378.1; -  
DR HSSP; P04475; 2BPR.  
DR InterPro; IPR001023; HSP70.  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PRO0301; HEATSHOCK70.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
DR Chaperone; ATP-binding; Heat shock.  
SQ SEQUENCE 641 AA; 69764 MW; 70FA5EBE144825CB CRC64;

Query Match	40.4%	Score 42;	DB 1;	Length 641;
Best Local	52.9%	Pred. No. 98;		
Matches 9;	Conservative 2;	Mismatches 6;	Indels 0;	Gaps 0;

```
QY      1 SMSIARLGKLA LAKK 17
        | : | | | : | | | |
Db      398 SLGIETLGGVMTKLKK 414
```

## RESULT 12

ID	TIRK_ECOLI	STANDARD;	PRT;	1188 AA.
----	------------	-----------	------	----------

DT 01-NOV-1988 (Rel. 09, Created)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

GN HSDR OR HSR OR B4350.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

NCBI\_TaxID=562;

RP SEQUENCE FROM N.A.

RX MEDLINE=88118919; PubMed=3323532;

RT "Organization and sequence of the hsd genes of *Escherichia coli*

RL J. Mol. Biol. 198:159-170(1987).

RP SEQUENCE FROM N.A.

RX MEDLINE=95334362; PubMed=7610040;

RA Blatner F.R.;

•

RT	"Analysis of the Escherichia coli genome VI: DNA sequence of the		
RT	region from 92.8 through 100 minutes.":		
RL	Nucleic Acids Res. 23:2105-2119(1995).		
RN	[3]		
RP	SEQUENCE OF 1-27 FROM N.A.		
RC	STRAIN-K12 / CR63;		
RX	MEDLINE=91317743; PubMed=1650347;		
RA	Walte-Rees P.A., Keating C.J., Moran L.S., Slatko B.E., Hornstra L.J.,		
RT	Benner J.S.;		
RT	"Characterization and expression of the Escherichia coli Mtr		
RL	restriction system.":		
RL	J. Bacteriol. 173:5207-5219(1991).		
CC	-1- FUNCTION: THE ECOK1 ENZYME RECOGNIZES 5'ACACN(6)GTGC-3'. SUBUNIT R		
CC	IS REQUIRED FOR BOTH NUCLEASE AND ATPASE ACTIVITIES, BUT NOT FOR		
CC	MODIFICATION.		
CC	-1- SUBUNIT: THE TYPE I RESTRICTION/MODIFICATION SYSTEM IS COMPOSED		
CC	OF THREE POLYPEPTIDES R,M AND S.		
CC	-1- MISCELLANEOUS: TYPE I RESTRICTION AND MODIFICATION ENZYMES ARE		
CC	COMPLEX, MULTIFUNCTIONAL SYSTEMS WHICH REQUIRE ATP, S-ADENOSYL		
CC	METHIONINE AND MG(2+) AS CO-FACTORS AND, IN ADDITION TO THEIR		
CC	ENDONUCLEOLYTIC AND METHYLASE ACTIVITIES, ARE POTENT DNA-DEPENDENT		
CC	ATPASES.		
CC	-1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1084		
CC	ONWARD AND IS SHORTER (1090 AA) DUE TO A FRAMESHIFT.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL; X06545; CA29791.1; ALT_FRAME.		
DR	EMBL; U14003; AAA97247.1; -		
DR	EMBL; AE000505; AAC77306.1; -		
DR	EMBL; X54198; CA38116.1; ALT_INIT.		
DR	PIR; Q00648; NDECKR.		
DR	REBASE; 980; ECOKI.		
DR	Ecogene; EG10459; hsdR.		
DR	InterPro: IPR001410; DEAD.		
DR	InterPro: IPR001650; Helicase_C.		
DR	Pfam; PF00271; helicase_C.1.		
DR	SMART; SM00487; DEXDC.1.		
DR	SMART; SM00490; HELICG.1.		
KW	Restriction system; Hydrolase; DNA-binding; ATP-binding;		
KW	Complete proteome.		
FT	DNA_BIND	449	468
FT	NP_BIND	490	496
FT	CONFLICT	629	697
FT	H-T-H MOTIF (BY SIMILARITY).		
FT	ATP (BY SIMILARITY).		
FT	DAVSIATLTALHTVQLTQEGFVRYRYRYAVIDGLFDQ		
FT	PIQIIITRNAQEGVLSKSGEYERISDP -> ECGKNSHR		
FT	HGGATVCADEFRRAGLPLTLYPYGRYRFRFSDRGASVSDHHQ		
FT	RAGGGLSIORRAGRAHO (IN REF. 1).		
SO	SEQUENCE	1188 AA;	136100 MW; CC0423F22A35E578 CR664;

Query Match	40.48;	Score 42;	DB 1;	Length 1188;
Best Local Similarity	57.98;	Pred. NO. 1.7e+02;		
Matches 11;	Conservative 3;	Mismatches 5;	Indels 0;	Gaps 0;

```
QY      5  ARLGKLA KLAKKLAKLAK 23
      |||  |:|:| |:|:|
Db     216  ARLALEAQLAEKNAELAK 234
```

## RESULT 13

POC_TOBAC	STANDARD;	PRT;	548 AA.
ID			

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE PROTOPORPHYRINOGEN OXIDASE, CHLOROPLAST PRECURSOR (EC 1.3.3.4) (PPO I)

```

DE (PROTOPORPHYRINOGEN IX OXIDASE ISOZYME I) (PPX I).
GN PPXI.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
ON NCBI_TaxID=4097;
RX MEDLINE=97385200; PubMed=9238074;
RA Lermontova I., Kruse E., Mock H.-P., Grimm B.;
RT "Cloning and characterization of a plastidial and a mitochondrial
RT isoform of tobacco protoporphyrinogen IX oxidase."
RT Proc. Natl. Acad. Sci. U.S.A. 94:8895-8900(1997).
CC -1- FUNCTION: CATALYZES THE 6-ELECTRON OXIDATION OF PROTOPORPHYRINOGEN
CC IX TO FORM PROTOPORPHYRIN IX.
CC -1- CATALYTIC ACTIVITY: PROTOPORPHYRINOGEN-IX + O(2) = PROTOPORPHYRIN-
CC IX + H(2)O(2).
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: PENULTIMATE STEP IN HEME AND PORPHYRIN BIOSYNTHESIS, AND
CC CHLOROPHYLL SYNTHESIS.
CC -1- SUBUNIT: HOMODIMER; CONTAINS ONE FAD PER HOMODIMER (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN EXPANDING PREATURE LEAVES.
CC DECREASED EXPRESSION IN OLDEST LEAVES. EXPRESSED AT VERY LOW LEVEL
CC IN ROOTS.
CC -1- INDUCTION: OSCILLATING EXPRESSION DURING DIURNAL GROWTH. MAXIMAL
CC EXPRESSION IN THE DARK PERIOD.
CC -1- SIMILARITY: BELONGS TO THE PROTOPORPHYRINOGEN OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y13465; CAA73865.1;
KW Protoporphyrinogen biosynthesis; Heme biosynthesis; Oxidoreductase;
KW Transist peptide; Flavoprotein; FAD; Chloroplast;
KW Chlorophyll biosynthesis.
FT TRANSIT 1 50 CHLOROPLAST (POTENTIAL).
FT CHAIN 51 548 PROTOPORPHYRINOGEN OXIDASE.
FT NP_BIND 78 83 FAD (POTENTIAL).
SQ SEQUENCE 548 AA; 59230 MW; 66892E78FB8A3E30 CRC64;

```

```

Query Match 39.9%; Score 41.5; DB 1; Length 548;
Best Local Similarity 57.9%; Pred. No. 1e+02;
Matches 11; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

```

```

OY 5 ARLGKTLAKTLAKTLAK 23
DB 303 ARLGSKL-KLSWKLSSITK 320

```

```

RESULT 14
HMGD_DROME STANDARD; PRT; 112 AA.
AC 005783; 09W2D3;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN D (HMG-D).
GN HMGD OR CG17950.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

```

RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 45-77.
RC TISSUE=Embryo;
RX MEDLINE=9223564; PubMed=1373803;
RA Wagner C.R., Hamana K., Elgin S.C.R.;
RT "A high-mobility-group protein and its cDNAs from Drosophila
RT melanogaster."
RT Mol. Cell. Biol. 12:1915-1923(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=94021387; PubMed=8414994;
RA Ner S.S., Churchill M.E.A., Seales M.A., Travers A.A.;
RT "HMG-Z, a second HMG-1-related protein in Drosophila melanogaster."
RT Nucleic Acids Res. 21:4369-4371(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Ashby A., An H.-J., Andrews-Planckoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fordin R.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kappen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=94222028; PubMed=8168480;
RA Ner S.S., Travers A.A.;
RT "HMG-D, the Drosophila melanogaster homologue of HMG I protein, is
RT associated with early embryonic chromatin in the absence of histone
RT H1."
RT EMBO J. 13:1817-1822(1994).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=95237208; PubMed=7720717;
RA Churchill M.E.A., Jones D.N.M., Glaser T., Hefner H., Seales M.A.,
RA Travers A.A.;
RT "HMG-D is an architecture-specific protein that preferentially binds
RT to DNA containing the dinucleotide TG."
RT EMBO J. 14:1264-1275(1995).

```

```

RN [6]
RN STRUCTURE BY NMR OF 2-74.
RX MEDLINE=95006330; Pubmed=7922039;
RA Jones D.N.M., Searles M.A., Shaw G.L., Churchill M.E.A., Ner S.S.,
RA Keeler J., Travers A.A., Neuhaus D.;
RT "The solution structure and dynamics of the DNA-binding domain of
RT HMG-D from Drosophila melanogaster.";
RL Structure 2:609-627(1994).
CC -1- FUNCTION: BINDS PREFERENTIALLY SINGLE-STRANDED DNA AND UNWINDS
CC DOUBLE STRANDED DNA. PREFERRED SITES CONTAINING THE SEQUENCE 5'-TTC-
CC 3. FACILITATES DNA BENDING. ASSOCIATED WITH EARLY EMBRYONIC
CC CHROMATIN IN THE ABSENCE OF HISTONE H1.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES OF DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE HMG1/HMG2 PROTEIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X71138; CAAS0468.1; -
DR EMBL: M77023; AAA28609.1; -
DR EMBL: AF003455; AAF46759.1; -
DR PIR: A44382; A44382.
DR PIR: S32724; S32724.
DR PDB: 1HMA; 3I-JUL-94.
DR FlyBase: FBgn0004362; HmgD.
DR InterPro: IPR000910; HMG_12_box.
DR Pfam: PF00505; HMG_box; 1.
DR SMART: SM00398; HMG; 1.
DR Nucleic Acid: Chromosomal protein; DNA-binding; 3D-structure.
KW DNA_BIND
FT DOMAIN 101 112 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 112 AA; 12416 MW; 3F537CCFD62FEC9F CRC64;

```

```

Query Match 39.4%; Score 41; DB 1; Length 112;
Best Local Similarity 56.5%; Pred. NO. 28;
Matches 13; Conservative 2; Mismatches 6; Indels 2; Gaps 1;
QY 3 SIARLGG-KLAKKAKKAKLAK 23
DB 77 SAANGGAKKRAKPAKVKAKSK 99

```

```

RESULT 15
Y458_METJA
ID Y458_METJA STANDARD: PRT: 216 AA.
AC 057900;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0458.
GN MJ0458.
OS Methanococcus jannaschlii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; Pubmed=8688087;
RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

```

```

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kleink H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "The complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67496; AAB98446.1; -
DR TIGR: MJ0458; -
DR InterPro: IPR001048; Aakkinase.
DR Pfam: PF00696; aakkinase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 216 AA; 24080 MW; 37856E333F3D2A76 CRC64;

```

```

Query Match 39.4%; Score 41; DB 1; Length 216;
Best Local Similarity 45.5%; Pred. NO. 50;
Matches 10; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
QY 2 MSARLGGKLAIAKAKKAKLAK 23
DB 1 MHIVKIGSLTYDAKPLKALK 22

```

Search completed: May 8, 2002, 13:35:58  
Job time: 110 sec



**THIS PAGE BLANK (USPTO)**

[illegible]

GN P115 OR U0140.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEKOVAR 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
RT urealyticum.";  
RL Nature 407:757-762(2000).  
DR EMBL: AE002114; AAF30546.1; -  
DR InterPro: IPR003439; ABC\_Transport.  
DR InterPro: IPR003405; SMC\_C.  
DR InterPro: IPR003395; SMC\_N.  
DR Pfam: PF02483; SMC\_C; 1.  
DR Pfam: PF02463; SMC\_N; 1.  
DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 981 AA; 111673 MW; 119100B1C9880104 CRC64;

Query Match 45.2%; Score 47; DB 2; Length 981;  
Best Local Similarity 52.4%; Pred. No. 2e+02;  
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 3 STARLGKTLAKKLAKLAK 23  
DB 188 NLARLNDIVANLKLAKLQK 208

RESULT 3  
O9KID1 PRELIMINARY; PRT; 364 AA.  
AC O9KID1;  
DT 01-OCT-2000 (TREMBLrel. 15; Created)  
DT 01-OCT-2000 (TREMBLrel. 15; Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17; Last annotation update)  
DE VLP21/25.  
GN VLP21/25.  
OS Borrelia hermsli.  
OC Bacterlia; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=140;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HSL;  
RA Barbour A.G., Carter C.J., Sohaskey C.D.;  
RT "Surface protein variation by expression site switching in the  
RT relapsing fever agent Borrelia hermsli.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF236049; AAF73949.1; -  
DR InterPro: IPR000680; Borrelia\_Lipo.  
DR Pfam: PF00921; Lipoprotein\_2; 1.  
KW Plasmid.  
SQ SEQUENCE 364 AA; 36716 MW; C5BF0DE4059EDB87 CRC64;

Query Match 43.3%; Score 45; DB 2; Length 364;  
Best Local Similarity 45.5%; Pred. No. 1.3e+02;  
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 SMSIRLGLKTLAKKLAKLA 22  
DB 270 ALKFARGGNGQLAKKEAKKA 291

RESULT 4  
O53164 PRELIMINARY; PRT; 475 AA.  
ID O53164

AC O53164;  
DT 01-NOV-1996 (TREMBLrel. 01; Created)  
DT 01-MAY-1999 (TREMBLrel. 10; Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17; Last annotation update)  
DE UPTAKE HYDROGENASE.  
GN HUPV OR HUPV2.  
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TaxID=1063;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RV;  
RA Franchi E., Tosi C., Rodriguez F., Pedroni P.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-64 FROM N.A.  
RC STRAIN=2.4.1;  
RX MEDLINE=96032007; PubMed=7551045;  
RA Gomelsky M., Kaplan S.;  
RT "Isolation of regulatory mutants in photosynthesis gene expression in  
RT Rhodobacter sphaeroides 2.4.1 and partial complementation of a prb  
RT mutant by the HupT histidine-kinase.";  
RL Microbiology 141:1805-1819(1995).  
DR EMBL: Y14197; CA74585.1; -  
DR EMBL: U37195; AA949491.1; -  
DR HSP; P21852; H2A.  
DR InterPro: IPR001501; Nifese\_Hases.  
DR Pfam: PF00374; Nifese\_Hases; 3.  
SQ SEQUENCE 475 AA; 50546 MW; 27686D37B7BCE758 CRC64;

Query Match 43.3%; Score 45; DB 2; Length 475;  
Best Local Similarity 45.0%; Pred. No. 1.7e+02;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 4 TARLGGKTLAKKLAKLAK 23  
DB 343 LAROGGVLRVGPRLLELAR 362

RESULT 5  
O9V1R8 PRELIMINARY; PRT; 1177 AA.  
AC O9V1R8;  
DT 01-MAY-2000 (TREMBLrel. 13; Created)  
DT 01-MAY-2000 (TREMBLrel. 13; Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17; Last annotation update)  
DE CHROMOSOME SEGREGATION PROTEIN (SMC1).  
GN PAB2109.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=29292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORSAY;  
RA Hellig R.;  
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
RT structure and evolution.";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ248284; CAB49281.1; -  
DR InterPro: IPR003439; ABC\_Transport.  
DR InterPro: IPR003405; SMC\_C.  
DR InterPro: IPR003395; SMC\_N.  
DR Pfam: PF02483; SMC\_C; 1.  
DR Pfam: PF02463; SMC\_N; 1.  
KW Complete proteome.  
SQ SEQUENCE 1177 AA; 135422 MW; 0DC17B1A18209F69 CRC64;

Query Match 43.3%; Score 45; DB 1; Length 1177;  
Best Local Similarity 50.0%; Pred. No. 4.6e+02;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;



OY 4 IARLGKLAIAKLAKLAK 23  
 DB 754 IAKLRGKIERLEKRRKDLAK 773

RESULT 6  
 O9N3T9 PRELIMINARY; PRT: 1262 AA.  
 AC O9N3T9; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
 DE Y47G6A.12 PROTEIN.  
 GN Y47G6A.12.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;  
 OC Rhabdilitidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=990613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Minx P., Graves T., Hawrysko C.;  
 RT "The sequence of C. elegans cosmid Y47G6A."  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC024791; AAF60651.1;  
 DR InterPro: IPR001064; CRYSTALLIN.  
 DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
 SQ SEQUENCE 1262 AA; 144120 MW; 4CBAB9E0330CEB3 CRC64;

Query Match 43.3%; Score 45; DB 5; Length 1262;  
 Best Local Similarity 50.0%; Pred. No. 4.9e+02;  
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 MSIARLGKLAIAKLAKLAK 23  
 DB 814 MSPGOLGPKAAELAIKIKHLK 835

RESULT 7  
 O9QJ6 PRELIMINARY; PRT: 68 AA.  
 AC O9QJ6; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
 DE ENVELOPE PROTEIN (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxId=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-120N;  
 RA Lopez-Galindez C.;  
 RT Genetic analysis of HIV-1 Spanish isolates."  
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF15816; AAF08454.1;  
 DR InterPro: IPR000777; GP120.

DR Pfam: PF00516; GP120; 1.  
 KM Envelope protein.  
 FT NON\_TER 1  
 FT NON\_TER 68  
 SQ SEQUENCE 68 AA; 7535 MW; A52C05ACCD93284B CRC64;

Query Match 42.3%; Score 44; DB 12; Length 68;  
 Best Local Similarity 55.6%; Pred. No. 30;  
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 SMSIARLGKLAIAKLAKL 18  
 DB 49 NVSRARLGDRLQGIARL 66

RESULT 8  
 P70906 PRELIMINARY; PRT: 360 AA.  
 AC P70906; 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
 DE VARIABLE MAJOR PROTEIN 18.  
 GN VMP18.  
 OS Borrelia hermsli.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxId=140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HSL; ATCC 35209;  
 RA Restrepo B.I., Carter C.J., Infante D., Barbour A.G.;  
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U52149; AAB1739.1;  
 DR InterPro: IPR000680; Borrelia\_lipo.  
 DR Pfam: PF00921; Lipoprotein\_2; 1.  
 SQ SEQUENCE 360 AA; 36906 MW; 86204DE08F4EAA38 CRC64;

Query Match 42.3%; Score 44; DB 2; Length 360;  
 Best Local Similarity 52.9%; Pred. No. 1.8e+02;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 7 LGGKLAIAKLAKLAKLAK 23  
 DB 94 LGGKLGASDELEVAR 110

RESULT 9  
 O9Z902 PRELIMINARY; PRT: 454 AA.  
 AC O9Z902; 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
 DE BIOTIN CARBOXYLASE.  
 GN ACCC OR CPN0182.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxId=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE=9920606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis".  
 RL Nat. Genet. 21:385-389 (1999).  
 DR EMBL: AE001604; AAD18335.1;  
 DR HSSP: P24182; IBNC.  
 DR InterPro: IPR000901; CPsase.  
 DR Pfam: PF00289; CPsase\_L\_chain; 1.  
 DR PROSITE: PS00867; CPsase\_2; UNKNOWN\_1.



AC 09JY43;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN NMB1755.  
 GN NMB1755.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 NCBI\_TaxID=491;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,  
 RA Cotton M.D., Ullrich T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58."  
 RL Science 287:1809-1815(2000).  
 DR EMBL: AE002525; AAF42096.1; -  
 DR TIGR: NMB1755; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 114 AA; 13184 MW; 75C791CD480842D4 CRC64;

Query Match 41.3%; Score 43; DB 2; Length 114;  
 Best Local Similarity 47.6%; Pred. No. 73;  
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 3 STARLGKLAIAKLAIAK 23  
 :|||:::|||||  
 Db 55 AIAIDGEMRLAKKSKMLK 75

RESULT 14  
 OY3548 PRELIMINARY; PRT; 246 AA.  
 AC 093548;  
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE F21G4.4 PROTEIN.  
 GN F21G4.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Mortimore B.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten T., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL Nature 368:32-38(1994).  
 DR EMBL: Z81016; CAB02663.1; -

DR InterPro: IPR001965; PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 SQ SEQUENCE 246 AA; 27784 MW; C3101C25B06046D CRC64;

Query Match 41.3%; Score 43; DB 5; Length 246;  
 Best Local Similarity 60.0%; Pred. No. 17e+02;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 9 GKLAIAKLAIAK 23  
 ||:::|||||  
 Db 149 GKIAKVDKTKAKTVK 163

RESULT 15  
 ID 09HT75 PRELIMINARY; PRT; 307 AA.  
 AC 09HT75;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE PROBABLE ADHESIN.  
 GN PA5498.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NCBI\_TaxID=287;  
 OX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mircoguchi S.D., Warriner P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004962; AAG08883.1; -  
 DR InterPro: IPR001987; Lipoprotein\_4.  
 DR Pfam: PF01297; Lipoprotein\_4; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 307 AA; 33608 MW; BE64F18B39B3BBA6 CRC64;

Query Match 41.3%; Score 43; DB 2; Length 307;  
 Best Local Similarity 58.8%; Pred. No. 2.1e+02;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 6 RLGGKLAIAKLAIA 22  
 |||||:::|||||  
 Db 180 RLGGLDGKLAERLCKLA 196

Search completed: May 8, 2002, 13:35:22  
 Job time: 190 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 13:33:37 ; Search time 12.86 seconds  
(without alignments)  
136.238 Million cell updates/sec

Title: CHIMERA  
Perfect score: 104  
Sequence: 1 SMSSTARLGKLAKLAKLAKLAK 23

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	45.2	981	C82930	pl15 protein UUI140
2	45	43.3	362	I40304	outer membrane lip
3	45	43.3	462	T17948	ABC transporter pr
4	45	43.3	1177	B71500	chromosome segrega
5	44	42.3	454	E72110	biotin carboxylase
6	44	42.3	454	F86513	biotin carboxylase
7	44	42.3	454	G81560	acetyl-coenzyme A
8	43.5	41.8	548	A81070	tspB protein, prob
9	43	41.3	104	H71555	hypothetical prote
10	43	41.3	114	B81046	hypothetical prote
11	43	41.3	246	T21215	hypothetical prote
12	43	41.3	307	C82959	probable adhesin p
13	43	41.3	337	C70390	hypothetical prote
14	43	41.3	412	D71372	hypothetical prote
15	43	41.3	755	S32103	filensin - bovine
16	43	41.3	790	JC5749	DNA topoisomerase
17	43	41.3	1722	I78879	retinoblastoma bin
18	42	40.4	247	A75254	conserved hypothet
19	42	40.4	273	F52M2	ribosomal protein
20	42	40.4	293	DDBF32	helix-destabilizin
21	42	40.4	293	DDBP36	helix-destabilizin
22	42	40.4	352	S16547	neutral proteinase
23	42	40.4	457	D71553	probable biotin ca
24	42	40.4	457	D81708	acetyl-coenzyme A
25	42	40.4	785	S63652	hypothetical prote
26	42	40.4	933	S75285	methyl-accepting c
27	42	40.4	1188	1 NDECKR	type I site-specif
28	42	40.4	1616	T47801	hypothetical prote
29	42	40.4	3534	T42567	tegument protein 2

30 41.5 39.9 548 2 T04058  
31 41 39.4 112 2 A44382  
32 41 39.4 140 2 A75187  
33 41 39.4 216 1 B64357  
34 41 39.4 221 2 F69526  
35 41 39.4 227 2 T15035  
36 41 39.4 240 2 A37004  
37 41 39.4 241 2 T27867  
38 41 39.4 242 2 A81782  
39 41 39.4 242 2 C81205  
40 41 39.4 262 2 F85616  
41 41 39.4 262 2 D64830  
42 41 39.4 293 2 B86133  
43 41 39.4 353 2 G96590  
44 41 39.4 354 2 T43507  
45 41 39.4 364 2 S11981

ALIGNMENTS

RESULT 1  
C82930  
pl15 protein UUI140 [imported] - Ureaplasma urealyticum  
C;Species: Ureaplasma urealyticum  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: C82930  
R;Glass, J.I.; Leikowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a  
A;Reference number: A82870  
A;Accession: C82930  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-981 <GLA>  
A;Cross-references: GB:AE002114; GB:AF222894; MID:G6899086; PIDN:AAF30546.1; GSPDB:GN  
A;Experimental source: serovar 3; biovar 1  
C;Genetics:  
A;Gene: pl15; UUI140  
A;Genetic code: SGC3

Query Match 45.2% Score 47; DB 2; Length 981;  
Best Local Similarity 52.4%; Pred. No. 71;  
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 SIARLGKGLAKLAKLAKLAK 23  
:||||:|:|||||  
DB 188 NLARLNDIVANLKKELAKLQK 208

RESULT 2  
I40304  
outer membrane lipoprotein - Borrelia hermsii  
C;Species: Borrelia hermsii  
C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 08-Oct-1999  
C;Accession: I40304  
R;Restrepo, B.I.; Kitten, T.; Carter, C.J.; Infante, D.; Barbour, A.G.  
Mol. Microbiol. 6, 3299-3311, 1992  
A;Title: Subtelomeric expression regions of Borrelia hermsii linear plasmids are high  
A;Reference number: I40300; MUID:93133110  
A;Accession: I40304  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-362 <RES>  
A;Cross-references: GB:L04787; MID:g144029; PIDN:AAA22965.1; PID:g144030  
C;Genetics:  
A;Gene: vmp25

Query Match 43.3% Score 45; DB 2; Length 362;  
Best Local Similarity 45.5%; Pred. No. 55;  
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 SMSIARLGKGLAKLAKLAKLA 22  
:: || || :||: ||  
Db 268 ALKFARGGGNAGOLAKERAKAA 289

RESULT 3  
Ti7948  
ABC transporter protein homolog A445L - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: Ti7948  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: Ti7948  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-462 <GRA>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96813.1  
A:Experimental source: specific host Chlorella strain NC64  
C:Genetics:  
A:Note: A445L

Query Match 43.3%; Score 45; DB 2; Length 462;  
Best Local Similarity 36.4%; Pred. NO. 68;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 2 MSIARLGKGLAKLAKLAKLA 23  
| : || :||: || :|  
Db 309 MGVIKSGGSVARLKAIPRIK 330

RESULT 4  
B75150  
chromosome segregation protein (smc1) PAB2109 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: B75150  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure  
A:Reference number: A75001  
A:Accession: B75150  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1177 <KAW>  
A:Cross-references: GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB49281.1; PID:g5457730  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB2109  
C:Superfamily: chromosome segregation protein SMC1

Query Match 43.3%; Score 45; DB 2; Length 1177;  
Best Local Similarity 50.0%; Pred. NO. 1.6e+02;  
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 IARLGKGLAKLAKLAKLAKLA 23  
||: |||: :| :| ||  
Db 754 IAKLRGKIERLERKRDKLKK 773

RESULT 5  
E72110  
biotin carboxylase - Chlamydomophila pneumoniae (strain CWL029)  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: E72110  
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606  
A:Accession: E72110  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <ARN>  
A:Cross-references: GB:AE001604; GB:AE001363; NID:g4376438; PIDN:AAD18335.1; PID:g4376437  
A:Experimental source: strain CWL029  
C:Genetics:  
C:Superfamily: biotin carboxylase; biotin carboxylase homology  
F;3-448/Domain: biotin carboxylase homology <BCH>

Query Match 42.3%; Score 44; DB 2; Length 454;  
Best Local Similarity 56.5%; Pred. NO. 93;  
Matches 13; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 SMSIARLGKGLAKLAKLAKLA 23  
| ||| :| :| || ||  
Db 106 SESIAMMGDKIA--AKSLAKTIK 126

RESULT 6  
F86513  
biotin carboxylase [imported] - Chlamydomophila pneumoniae (strain J138)  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: F86513  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: F86513  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <STO>  
A:Cross-references: GB:BA000008; NID:g8978555; PIDN:BAA98392.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: accC  
C:Superfamily: biotin carboxylase; biotin carboxylase homology

Query Match 42.3%; Score 44; DB 2; Length 454;  
Best Local Similarity 56.5%; Pred. NO. 93;  
Matches 13; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 SMSIARLGKGLAKLAKLAKLA 23  
| ||| :| :| || ||  
Db 106 SESIAMMGDKIA--AKSLAKTIK 126

RESULT 7  
G81560  
acetyl-coenzyme A carboxylase, biotin carboxylase CP0586 [imported] - Chlamydomophila p  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: G81560  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, W.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: G81560  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <REA>  
A:Cross-references: GB:AE002217; GB:AE002161; NID:g7189493; PIDN:AAF38404.1; PID:g7189493  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0586  
C:Superfamily: biotin carboxylase; biotin carboxylase homology

RESULT 10

C82959  
probable adhesin PA5498 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C82959  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey,  
J.M.; Landman, R.; Tarran, P.; Brody, L.L.; Coulter, K.R.; Kas, A.; Larson,  
M.S.; Yuan, Y.; Brody, L.L.; Olson, M.V.  
.: Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportu-

A:Reference number: A82950; MUID:20437337  
A:Accession: C82959  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 <STO>  
A:Cross-references: GB:AE004962; GB:AE004091; NID:g9951826; PIDN:AAG08883.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA5498  
C:Superfamily: hypothetical protein HI0119

Query Match 41.3%; Score 43; DB 2; Length 307;  
Best Local Similarity 58.8%; Pred. No. 90;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 RLGGKLA LAKKLA 22  
||||| ||:|:|  
DB 180 RLGGDGLRLRLGKLA 196

RESULT 13  
C70990  
hypothetical protein Rv3128c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 24-Jul-1998  
C:Accession: C70990  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: C70990  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-337 <COL>  
A:Cross-references: GB:Z95150; GB:AL123456; NID:g3250708  
A:Experimental source: strain H37RV  
A:Note: This ORF is annotated but not translated in GenBank entry M7CY164, release 106  
C:Genetics:  
A:Gene: Rv3128c  
A:Start codon: GTG  
F:194/Region: amber stop codon

Query Match 41.3%; Score 43; DB 2; Length 337;  
Best Local Similarity 43.5%; Pred. No. 97;  
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 SMSIARLGK LAKKLA 23  
|::|::|::|::|  
DB 288 SLNPAQIGRKRIADLQNRLLILAK 310

RESULT 14  
D71972  
hypothetical protein jhp0110 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Jun-2000  
C:Accession: D71972  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A:Reference number: A71800; MUID:99120557  
A:Accession: D71972  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-412 <ARN>  
A:Cross-references: GB:AE001450; GB:AE001439; NID:g4154617; PIDN:AAD05689.1; PID:g415461

A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0110  
C:Superfamily: Helicobacter pylori hypothetical protein HP0120

Query Match 41.3%; Score 43; DB 2; Length 412;  
Best Local Similarity 38.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 3 SIARLGK LAKKLA 23  
:::|:|:|:|:|  
DB 38 TLRSTGGITRIAKKRELK 58

RESULT 15  
S32103  
filensin - bovine  
N:Alternate names: Intermediate filament protein  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 10-Sep-1997  
C:Accession: S32103; A40690  
R:Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; FitzGerald, P.G.; Ouzounis, C.; Geo  
submitted to the EMBL Data Library, March 1993  
A:Description: Bovine filensin possesses primary and secondary structure similarity t  
A:Reference number: S32103  
A:Accession: S32103  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-755 <GO>  
A:Cross-references: EMBL:X7388; NID:g287751; PID:g287752  
R:Gounari, F.; Merdes, A.; Quinlan, R.; Hess, J.; FitzGerald, P.G.; Ouzounis, C.A.; G  
J. Cell Biol. 121, 847-853, 1993.  
A:Title: Bovine filensin possesses primary and secondary structure similarity to inte  
A:Reference number: A40690; MUID:93260017  
A:Accession: A40690  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-622, 'RP', 625-755 <GO>  
A:Experimental source: lens  
A:Note: sequence extracted from NCBI backbone (NCBIN:132495, NCBIp:132499)  
A:Note: part of this sequence was confirmed by protein sequencing  
C:Keywords: membrane-associated protein

Query Match 41.3%; Score 43; DB 2; Length 755;  
Best Local Similarity 50.0%; Pred. No. 2e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 GSKLAK LAKKLA 23  
|::|:|:|:|:|  
DB 431 GSKSKAFKLGKMK 446

Search completed: May 8, 2002, 13:35:41  
Job time: 124 sec





**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	72	69.2	28	1	US-08-944-133-7	Sequence 7, Appli
2	72	69.2	21	1	US-08-944-133-4	Sequence 4, Appli
3	72	69.2	28	1	US-08-944-133-8	Sequence 8, Appli
4	72	69.2	28	1	US-08-944-133-27	Sequence 27, Appl
5	67	64.4	21	1	US-08-944-133-26	Sequence 26, Appl
6	66	63.5	1	1	US-08-944-133-3	Sequence 3, Appli
7	62	59.6	14	1	US-08-944-133-6	Sequence 6, Appli
8	62	59.6	21	1	US-07-908-455A-4	Sequence 4, Appli
9	62	59.6	21	1	US-08-434-120-30	Sequence 30, Appl
10	62	59.6	21	1	US-08-465-325-30	Sequence 30, Appl
11	62	59.6	23	2	US-08-723-306-27	Sequence 27, Appl
12	62	59.6	23	5	PCR-US96-10041-27	Sequence 27, Appl
13	61	58.7	23	2	US-08-723-306-28	Sequence 28, Appl
14	61	58.7	23	5	PCR-US96-10041-28	Sequence 28, Appl
15	60	57.7	21	1	US-07-908-455A-2	Sequence 2, Appli
16	60	57.7	21	1	US-07-908-455A-28	Sequence 28, Appl
17	60	57.7	21	1	US-08-434-120-28	Sequence 28, Appl
18	60	57.7	21	1	US-08-434-120-54	Sequence 54, Appl
19	60	57.7	21	1	US-08-465-325-28	Sequence 28, Appl
20	60	57.7	21	1	US-08-465-325-54	Sequence 54, Appl
21	60	57.7	24	1	US-07-908-455A-43	Sequence 43, Appl
22	60	57.7	24	1	US-08-434-120-69	Sequence 69, Appl
23	58	55.8	28	1	US-08-944-133-38	Sequence 38, Appl
24	58	55.8	28	1	US-08-944-133-22	Sequence 22, Appl
25	58	55.8	28	1	US-08-944-133-39	Sequence 39, Appl
26	58	55.8	28	1	US-08-944-133-43	Sequence 43, Appl
27	57	54.8	14	1	US-08-944-133-25	Sequence 25, Appl

Best Local Similarity 80.0%; Pred. No. 0.00025;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGGKLA LAKLAKLAK 23  
DB 2 LAKLAKLAKLAKLAKLAK 21

## RESULT 2

US-08-944-133-4  
; Sequence 4, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Runnels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944.133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Runnels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-4

Query Match 69.2%; Score 72; DB 1; Length 28;  
Best Local Similarity 80.0%; Pred. No. 0.00033;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGGKLA LAKLAKLAK 23  
DB 6 LAKLAKLAKLAKLAKLAK 25

## RESULT 3

US-08-944-133-8  
; Sequence 8, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L

; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Runnels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944.133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Runnels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-8

Query Match 69.2%; Score 72; DB 1; Length 28;  
Best Local Similarity 80.0%; Pred. No. 0.00033;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGGKLA LAKLAKLAK 23  
DB 2 LAKLAKLAKLAKLAKLAK 21

## RESULT 4

US-08-944-133-27  
; Sequence 27, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Runnels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:



QY 6 RLGGKLAKLAKLAKLAK 23  
:| | | | | | | | | |  
Db 1 KLAKLAKLAKLAKLAKLAK 18

RESULT 7  
US-08-944-133-6  
; Sequence 6, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Runnels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944.133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Runnels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-6

Query Match 59.6%; Score 62; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.0044;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 KLAKLAKLAKLAKLAK 23  
| | | | | | | | | |  
Db 1 KLAKLAKLAKLAKLAK 14

RESULT 8  
US-07-908-455A-4  
; Sequence 4, Application US/07908455A  
; Patent No. 5459237  
; GENERAL INFORMATION:  
; APPLICANT: Berkowitz, Barry A.  
; APPLICANT: Kari, U. Prasad  
; APPLICANT: Maloy, W. Lee  
; TITLE OF INVENTION: No. 5459237el Peptide Compositions and

; TITLE OF INVENTION: Uses Therefor  
; NUMBER OF SEQUENCES: 89  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,  
; ADDRESSEE: Cecchi & Stewart  
; STREET: 6 Becker Farm Road  
; CITY: Roseland  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: DW4.V2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/908.455A  
; FILING DATE: 19920702  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07686115  
; FILING DATE: 15-APR-1991  
; APPLICATION NUMBER: US 07476629  
; FILING DATE: 08-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Olstein, Elliot M.  
; REGISTRATION NUMBER: 24,025  
; REFERENCE/DOCKET NUMBER: 421250-122  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; OTHER INFORMATION: amide-terminated  
US-07-908-455A-4

Query Match 59.6%; Score 62; DB 1; Length 21;  
Best Local Similarity 82.4%; Pred. No. 0.0064;  
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 RLGGKLAKLAKLAKLA 22  
:| | | | | | | | | |  
Db 1 KLAKLAKLAKLAKLA 17

RESULT 9  
US-08-434-120-30  
; Sequence 30, Application US/08434120  
; Patent No. 5635479  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Margaret A.  
; APPLICANT: Jacob, Leonard S.  
; APPLICANT: Maloy, W. Lee  
; TITLE OF INVENTION: Treatment of Gynecological  
; TITLE OF INVENTION: Malignancies with  
; TITLE OF INVENTION: Biologically Active Peptides  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,  
; ADDRESSEE: Cecchi & Stewart  
; STREET: 6 Becker Farm Road  
; CITY: Roseland  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:

;/ MEDIUM TYPE: 3.5 inch diskette  
;/ COMPUTER: IBM PS/2  
;/ OPERATING SYSTEM: PC-DOS  
;/ SOFTWARE: DW4.V2  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/434,120  
;/ FILING DATE:  
;/ CLASSIFICATION: 514  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/297,950  
;/ FILING DATE:  
;/ APPLICATION NUMBER: US/08/226,108  
;/ FILING DATE:  
;/ APPLICATION NUMBER: US/07/937,462  
;/ FILING DATE:  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Olstein, Elliot M.  
;/ REGISTRATION NUMBER: 24,025  
;/ REFERENCE/DOCKET NUMBER: 421250-194  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 201-994-1700  
;/ TELEFAX: 201-994-1744  
;/ INFORMATION FOR SEQ ID NO: 30:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 21 amino acids  
;/ TYPE: amino acid  
;/ STRANDEDNESS:  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: peptide  
;/ US-08-434-120-30

Query Match 59.6%; Score 62; DB 1; Length 21;  
Best Local Similarity 82.4%; Pred. No. 0.0064; 2; Indels 0;  
Matches 14; Conservative 1; Mismatches 2; Gaps 0;

QY 6 RLGKGLAKLAKLAKLA 22  
: | | | | | | | | | |  
Db 1 KLAGKLAKLAGKLAKLA 17

RESULT 10  
US-08-465-325-30  
; Sequence 30, Application US/08465325  
; Patent No. 5686563  
; GENERAL INFORMATION:  
; APPLICANT: Magainin Pharmaceuticals Inc.  
; APPLICANT: 5110 Campus Drive  
; APPLICANT: Plymouth Meeting, PA 19462  
; TITLE OF INVENTION: Biologically Active Peptides Having  
; TITLE OF INVENTION: N-Terminal Substitutions  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I. Street, N.W. Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,325  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/184,462  
; FILING DATE: 18-JAN-94  
; PRIOR APPLICATION DATA:

;/ APPLICATION NUMBER: 07/891,201  
;/ FILING DATE: 01-JUN-92  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Fordis, Jean B  
;/ REGISTRATION NUMBER: 32,984  
;/ REFERENCE/DOCKET NUMBER: 05387.0021-03000  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (202) 408-4000  
;/ TELEFAX: (202) 408-4400  
;/ INFORMATION FOR SEQ ID NO: 30:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 21 amino acids  
;/ TYPE: amino acid  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: peptide  
;/ US-08-465-325-30

Query Match 59.6%; Score 62; DB 1; Length 21;  
Best Local Similarity 82.4%; Pred. No. 0.0064; 2; Indels 0;  
Matches 14; Conservative 1; Mismatches 2; Gaps 0;

QY 6 RLGKGLAKLAKLAKLA 22  
: | | | | | | | | | |  
Db 1 KLAGKLAKLAGKLAKLA 17

RESULT 11  
US-08-723-306-27  
; Sequence 27, Application US/08723306  
; Patent No. 5856178  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth  
; APPLICANT: Morrey PhD, John  
; APPLICANT: Reed, William  
; TITLE OF INVENTION: Cassette for Expression of Lytic  
; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,306  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert PhD, Susan E  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2549  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 8015321922  
; TELEFAX: 8015319168  
;/ INFORMATION FOR SEQ ID NO: 27:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 23 amino acids  
;/ TYPE: amino acid  
;/ STRANDEDNESS:  
;/ TOPOLOGY: not relevant  
;/ MOLECULE TYPE: peptide  
;/ HYPOTHETICAL: YES  
;/ US-08-723-306-27

Query Match 59.6%; Score 62; DB 2; Length 23;

Best Local Similarity 70.0%; Pred. No. 0.007;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 IARLGGKLA LAKKLAKLAK 23  
: : | | | | | | | | | |  
DB 2 LKLA LKLA LKLA LKLA LKLA 21

## RESULT 12

PCT-US96-10041-27  
; Sequence 27, Application PC/TUS9610041  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth  
; APPLICANT: Morrey PhD, John  
; APPLICANT: Reed, William  
; TITLE OF INVENTION: Cassette for Expression of Lytic  
; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/10041  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert PhD, Susan E  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2549  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 8015321922  
; TELEFAX: 8015319168  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: YES  
PCT-US96-10041-27

Query Match 59.6%; Score 62; DB 5; Length 23;  
Best Local Similarity 70.0%; Pred. No. 0.007;  
Matches 14; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 IARLGGKLA LAKKLAKLAK 23  
: : | | | | | | | | | |  
DB 2 LKLA LKLA LKLA LKLA LKLA 21

## RESULT 13

US-08-723-306-28  
; Sequence 28, Application US/08723306  
; Patent No 5856178  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth  
; APPLICANT: Morrey PhD, John  
; APPLICANT: Reed, William  
; TITLE OF INVENTION: Cassette for Expression of Lytic  
; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Trask Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,306  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert PhD, Susan E  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2549  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 8015321922  
; TELEFAX: 8015319168  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: YES  
US-08-723-306-28

Query Match 58.7%; Score 61; DB 2; Length 23;  
Best Local Similarity 77.8%; Pred. No. 0.0097;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 RLGGKLA LAKKLAKLAK 23  
: | | | | | | | | | |  
DB 4 LKLA LKLA LKLA LKLA LKLA 21

## RESULT 14

PCT-US96-10041-28  
; Sequence 28, Application PC/TUS9610041  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth  
; APPLICANT: Morrey PhD, John  
; APPLICANT: Reed, William  
; TITLE OF INVENTION: Cassette for Expression of Lytic  
; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/10041  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert PhD, Susan E  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2549  
; TELECOMMUNICATION INFORMATION:



TELEPHONE: 8015321922  
TELEFAX: 8015319168  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES  
PCT-US96-10041-28

Query Match 58.7%; Score 61; DB 5; Length 23;  
Best Local Similarity 77.8%; Pred. No. 0.0097;  
Matches 14; Conservative 1; Mismatches 3; Indels

```

RESULT 15
US-07-908-455A-2
; Sequence 2, Application US/07908455A
; Patent No. 5459237
; GENERAL INFORMATION:
; APPLICANT: Berkowitz, Barry A.
; APPLICANT: Karl, U. Prasad
; APPLICANT: Matoy, W. Lee
; TITLE OF INVENTION: No. 5459237e1 Peptide Compositions and
; TITLE OF INVENTION: Uses Thereof
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi & Stewart
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068

```

Query Match 57.7%; Score 60; DB 1; Length 21;  
Best Local Similarity 57.9%; Pred. No. 0.012;  
Matches 11; Conservative 6; Mismatches 2; Indels

Search completed: May 8, 2002, 13:34:22  
Job time: 550 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:20:14 ; Search time 35.89 seconds  
(without alignments)  
28.529 Million cell updates/sec

Title: US-09-765-086-207  
Perfect score: 30  
Sequence: 1 SMSIARL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_17.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	90.0	301	2 P77850	P77850 chlamydia t
2	27	90.0	301	2 O84705	O84705 chlamydia t
3	27	90.0	301	2 Q9PLM7	Q9PLM7 chlamydia m
4	27	90.0	329	10 Q9CGF0	Q9CGF0 arabidopsis
5	27	90.0	369	10 O65663	O65663 arabidopsis
6	26	86.7	305	1 Q9YCT2	Q9YCT2 aeropyrum p
7	26	86.7	307	2 O68069	O68069 rhodobacter
8	26	86.7	317	11 Q9CSE0	Q9CSE0 mus musculus
9	26	86.7	422	13 Q90417	Q90417 brachydanio
10	26	86.7	571	4 Q9BU29	Q9BU29 homo sapien
11	26	86.7	849	3 Q12205	Q12205 saccharomyc
12	26	86.7	1204	4 Q9H0H0	Q9H0H0 homo sapien
13	26	86.7	1209	4 Q9ULD3	Q9ULD3 homo sapien
14	26	86.7	1216	10 Q9SYI4	Q9SYI4 arabidopsis
15	26	86.7	3222	12 O85057	O85057 peanut stri
16	26	86.7	3222	12 O85071	O85071 peanut stri
17	26	86.7	3475	5 Q9UL75	Q9UL75 leishmania
18	25	83.3	212	5 Q97096	Q97096 clonorchis
19	25	83.3	213	10 Q9ZWB5	Q9ZWB5 arabidopsis

013888 schizosacch  
P92175 drosophila  
P91933 drosophila  
P91934 drosophila  
P91935 drosophila  
Q91128 pseudomonas  
Q55529 synechocyst  
Q9m2b7 arabidopsis  
Q9rj38 streptomyces  
O64427 chlorella v  
Q9emq2 ansacta moo  
Q9vih3 drosophila  
O48436 bacterioph  
Q91195 pseudomonas  
Q52607 agrobacteri  
O64748 avian adeno  
O87623 pseudomonas  
Q9zma6 helicobacte  
O25100 helicobacte  
Q9fc82 streptomyce  
O17162 caenorhabdi  
Q9km47 vibrio chol  
Q9hlj4 thermoplas  
Q9yaq7 aeropyrum p  
Q9kfl4 bacillus ha  
O81487 arabidopsis

#### ALIGNMENTS

RESULT 1  
ID P77850 PRELIMINARY; PRT; 301 AA.  
AC P77850; 1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE PHOSPHATIDYL SERINE DECARBOXYLASE (EC 4.1.1.65).  
GN PSD.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L2/434/BU;  
RA McClarty G.A.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYL-L-SERINE -  
CC PHOSPHATIDYLETHANOLAMINE + CO(2).  
CC -!- COFACTOR: PYRIDOXAL-PHOSPHATE OR PYRUVATE.  
DR EMBL: U72715; AAB17564.1; -;  
DR InterPro: IPR003817; PS\_Dcarboxylase.  
DR Pfam: PF02666; PS\_Dcarboxylase; 1.  
SQ SEQUENCE 301 AA; 34064 MW; A4E852492F408207 CRC64;

Query Match 90.08; Score 27; DB 2; Length 301;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
Db 158 SMSIARL 164

RESULT 2  
ID O84705 PRELIMINARY; PRT; 301 AA.  
AC O84705;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PHOSPHATIDYLSERINE DECARBOXYLASE.  
 GN PSSD OR CT699.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D/UW-3/CX;  
 RX MEDLINE=99000809; PubMed=9784136;  
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis.";  
 RL Science 282:754-759(1998).  
 DR EMBL; AF001340; AAC68294.1; -;  
 DR InterPro; IPR003817; PS\_Dcarbxyase.  
 DR Pfam; PF02666; PS\_Dcarbxyase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 301 AA; 34076 MW; D82A7E0272CC896 CRC64;

Query Match 90.0%; Score 27; DB 2; Length 301;  
 Best Local Similarity 85.7%; Pred. No. 56;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 II:||||  
 Db 158 SMAILR 164

RESULT 3  
 Q9PLM7  
 ID Q9PLM7 PRELIMINARY; PRT; 301 AA.  
 AC Q9PLM7  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE PHOSPHATIDYLSERINE DECARBOXYLASE PROENZYME.  
 GN TC0072.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MOPN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AF002274; AAF38954.1; -;  
 DR TIGR; TC0072; -;  
 DR InterPro; IPR003817; PS\_Dcarbxyase.  
 DR Pfam; PF02666; PS\_Dcarbxyase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 301 AA; 34113 MW; 09E427CA5CD5ACF2 CRC64;

Query Match 90.0%; Score 27; DB 2; Length 301;  
 Best Local Similarity 85.7%; Pred. No. 56;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 II:||||  
 Db 158 SMAILR 164

RESULT 4  
 Q9C6F0  
 ID Q9C6F0 PRELIMINARY; PRT; 329 AA.  
 AC Q9C6F0;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYOSYAMINE 6-DIOXYGENASE HYDROXYLASE, PUTATIVE.  
 GN T32G9.27.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Arafujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,  
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).  
 DR EMBL; AC079605; AAG50602.1; -;  
 DR InterPro; IPR002419; Fe\_asc\_oxidored.  
 DR InterPro; IPR002283; IPN\_synth.  
 DR Pfam; PF00671; Fe\_Asc\_oxidored; 1.  
 DR PRINTS; PR00682; IPNSYNTHASE.  
 DR DIOXYGENASE.  
 KW DIOXYGENASE.  
 SQ SEQUENCE 329 AA; 37635 MW; C60C1DBCDB402747 CRC64;

Query Match 90.0%; Score 27; DB 10; Length 329;  
 Best Local Similarity 85.7%; Pred. No. 61;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 II:||||  
 Db 150 SMAILR 156

RESULT 5  
 O65663  
 ID O65663 PRELIMINARY; PRT; 369 AA.  
 AC O65663;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOTHETICAL 42.7 KDA PROTEIN.  
 GN T19P19.150 OR AT4G39760.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Monfort A., Casacuberta E., Puigdomenech P., Hohseisel J.,  
 RA Mewes H.W., Mayer K.F.X., Schueller C.;

RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RA Monfort A., Casacuberta E., Puigdomenech P., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.,  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL022605; CAA18762.1; -;  
 DR EMBL; AL161595; CAB80639.1; -;  
 DR Mendel; 29837; Arath:3427; 29837.  
 DR InterPro; IPR001810; F-box.  
 DR InterPro; IPR001798; Kelch.  
 DR Pfam; PF00646; F-box; 1.  
 DR Pfam; PF01344; Kelch; 1.  
 DR SMART; SM00256; FBOX; 1.  
 DR PROSITE; PS0181; FBOX; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 369 AA; 42660 MW; 900E2D1C01539684 CRC64;

Query Match 90.0%; Score 27; DB 10; Length 369;  
 Best Local Similarity 71.4%; Pred. No. 68;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 Db 162 SMSVARI 168

RESULT 6  
 Q9YCT2 PRELIMINARY; PRT; 305 AA.  
 AC Q9YCT2;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE 305AA LONG HYPOTHETICAL DTDp-4-DEHYDRORHAMNOSE REDUCTASE.  
 GN APE1179.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;  
 OC Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-po K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 crenarchaeon, Aeropyrum pernix K1";  
 RL DNA Res. 6:83-101(1999).  
 DR EMBL; AF000061; BAA80165.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 305 AA; 34301 MW; 33261BC74451923A CRC64;

Query Match 86.7%; Score 26; DB 1; Length 305;  
 Best Local Similarity 71.4%; Pred. No. 1e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 Db 99 SMSVARI 105

RESULT 7

O68069  
 ID O68069 PRELIMINARY; PRT; 307 AA.  
 AC O68069;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 33.2 KDA PROTEIN.  
 OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 OC Rhodobacter.  
 OX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SB1003;  
 RX MEDLINE=97404404; PubMed=9256491;  
 RA Vleck C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;  
 RT "sequence of a 189-kb segment of the chromosome of Rhodobacter  
 capsulatus SB1003";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
 DR EMBL; AF010496; AAC16155.1; -;  
 DR InterPro; IPR002606; FAD\_Synth.  
 DR Pfam; PF01687; FAD\_Synth; 1.  
 DR ProDom; PD003662; FAD\_Synth; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 307 AA; 33190 MW; 33CDFFED0FBDF66 CRC64;

Query Match 86.7%; Score 26; DB 2; Length 307;  
 Best Local Similarity 71.4%; Pred. No. 1e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 Db 205 NMSVARL 211

RESULT 8  
 Q9CSE0 PRELIMINARY; PRT; 317 AA.  
 AC Q9CSE0;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 2810417D08RIK PROTEIN (FRAGMENT).  
 GN 2810417D08RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakawa I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Ljungdahl S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontecki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).

```
DR EMBL; AK013101; BAB28647.1; -.
DR MGD; MGI:1917672; 2810417D08Rik.
FT NON_TER 1
SQ SEQUENCE 317 AA; 34869 MW; 7D99ADD59C36BC82 CRC64;

Query Match 86.7%; Score 26; DB 11; Length 317;
Best Local Similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 186 SLSVARL 192

RESULT 9
Q90417 PRELIMINARY; PRT; 422 AA.
AC Q90417;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE RETINOID X RECEPTOR DELTA.
GN RXRD OR RXR.
OS Brachydanio rerio (Zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009547; PubMed=7565671;
RA Jones B.B., Ohno C.K., Allenby G., Boffa M.B., Levin A.A.,
RA Grippo J.F., Petkovich M.;
RT "New retinoid X receptor subtypes in zebra fish (Danio rerio)
RT differentially modulate transcription and do not bind 9-cis retinoic
RT acid.";
RL Mol. Cell. Biol. 15:5226-5234(1995).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL; U29941; AAC59721.1; -.
DR HSP; P19793; 2NLL.
DR ZFIN; ZDB-GENE-990415-242; rxrd.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; znF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 422 AA; 46732 MW; 6CCDBA3CE29C7238 CRC64;

Query Match 86.7%; Score 26; DB 13; Length 422;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 53 SMSVSRL 59

RESULT 10
Q9BU29 PRELIMINARY; PRT; 571 AA.
AC Q9BU29;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DR EMBL; AK013101; BAB28647.1; -.
DR MGD; MGI:1917672; 2810417D08Rik.
FT NON_TER 1
SQ SEQUENCE 317 AA; 34869 MW; 7D99ADD59C36BC82 CRC64;

Query Match 86.7%; Score 26; DB 4; Length 571;
Best Local Similarity 71.4%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 519 SMSIARL 525

RESULT 11
Q12205 PRELIMINARY; PRT; 849 AA.
AC Q12205;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ORF YLR057W.
GN YLR057W OR L2153.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Urrestarazu L.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Andre B., Urrestarazu L.A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X94507; CAA64304.1; -.
DR EMBL; Z73229; CAA97587.1; -.
DR EMBL; Z73230; CAA97589.1; -.
DR SGD; S0004047; YLR057W.
DR InterPro; IPR001382; Glyco_hydro_47.
DR Pfam; PF01532; Glyco_hydro_47; 1.
DR ProDom; PD003239; Glyco_hydro_47; 1.
SQ SEQUENCE 849 AA; 96996 MW; A6B87AC32936A0D5 CRC64;

Query Match 86.7%; Score 26; DB 3; Length 849;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSIARL 7
Db 1 MSIARL 6

RESULT 12
Q9HOH0 PRELIMINARY; PRT; 1204 AA.
ID Q9HOH0
AC Q9HOH0;
```

```

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHEMETICAL 134.3 KDA PROTEIN.
GN DFXZP434H1220.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: ALI36800; CAB66734.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1204 AA; 134345 MW; 50B87F6FF2A1088 CRC64;

Query Match      86.7%; Score 26; DB 4; Length 1204;
Best Local Similarity 71.4%; Pred. No. 4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 1073 SLSVARL 1079

RESULT 13
Q9ULD3
ID Q9ULD3 PRELIMINARY; PRT; 1209 AA.
AC Q9ULD3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE KIAA1287 PROTEIN (FRAGMENT).
GN KIAA1287.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirotsava M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL: AB033113; BAA86601.1; -.
FT NON_TER 1 1
SQ SEQUENCE 1209 AA; 134941 MW; BF44BD99F71A37B9 CRC64;

Query Match      86.7%; Score 26; DB 4; Length 1209;
Best Local Similarity 71.4%; Pred. No. 4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 1078 SLSVARL 1084

RESULT 14
Q9SYI4
ID Q9SYI4 PRELIMINARY; PRT; 1216 AA.
AC Q9SYI4
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHEMETICAL 134.3 KDA PROTEIN.
GN T7B11.12 OR AT4G01860.

```

```

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Huang E.N., Nascimento L., de la Bastide M., Habermann K., Vil M.D.,
RA Preston R.R., Spiegel L.A., See L.H., Shah R., Matero A., Schutz K.,
RA O'Shaughnessy A., Rodriguez M., Shekher M., Swaby I., Schutz K.,
RA Parnell L.D., Dedhia N.N., McCombie W.R.;
RT "Arabidopsis thaliana BAC T7B11 from chromosome IV near 10 cm.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007138; AAD22648.1; -.
DR EMBL: AL161493; CAB80679.1; -.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_3.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 1216 AA; 134272 MW; C872A47A6666C0D7 CRC64;

Query Match      86.7%; Score 26; DB 10; Length 1216;
Best Local Similarity 71.4%; Pred. No. 4.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 601 SLSVARL 607

RESULT 15
Q85057
ID Q85057 PRELIMINARY; PRT; 3222 AA.
AC Q85057
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POLYPROTEIN.
OS Peanut stripe virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=28353;
RN [1]
RP SEQUENCE OF 2854-3222 FROM N.A.
RC STRAIN=BL0TCH;
RX MEDLINE=93168029; PubMed=7916587;
RA Cassidy B., Sherwood J.L., Nelson R.S.;
RT "Cloning of the capsid protein gene from a blotch isolate of peanut
RT stripe virus.";
RL Arch. Virol. 128:287-297(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BL0TCH;
RX MEDLINE=94358761; PubMed=8077957;
RA Gunesinghe U.B., Flisinski S., Nelson R.S., Cassidy B.G.;
RT "Nucleotide sequence and genome organization of peanut stripe
RT potyvirus.";

```

```

RL J. Gen. Virol. 75:2519-2525(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BLOTCH;
RA Guensinghe U.B.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.
DR EMBL; U05771; AA01025.1; -
DR MEROPS; C04.001; -
DR MEROPS; C06.001; -
DR MEROPS; S30.001; -
DR InterPro; IPR001917; Aminotransf_2.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001730; Peptidase_C4.
DR InterPro; IPR001456; Peptidase_C6.
DR InterPro; IPR001592; Poty_coat.
DR InterPro; IPR002340; Poty_P1.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00863; Peptidase_C4; 1.
DR Pfam; PF00851; Peptidase_C6; 1.
DR Pfam; PF00767; Poty_coat; 1.
DR Pfam; PF01577; Poty_P1; 1.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR PRINTS; PR00966; NIAPOTYPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
KW ATP-binding; Helicase; Polypeptide.
FT CHAIN 1 443 P1 PROTEIN.
FT CHAIN 444 900 HC-PRO.
FT CHAIN 901 1247 P3.
FT CHAIN 1248 1299 6 KI.
FT CHAIN 1300 1333 CI.
FT CHAIN 1934 1986 6K2.
FT CHAIN 1987 2176 NIA-VPG.
FT CHAIN 2177 2419 NIA-PRO.
FT CHAIN 2420 2935 NIB.
FT CHAIN 2936 3222 CP.
SQ SEQUENCE 3222 AA; 365723 MW; 96E3CE717D634B50 CRC64;

```

```

Query Match 86.7%; Score 26; DB 12; Length 3222;
Best Local Similarity 85.7%; Pred. No. 1.1e+03;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7
DB 558 SMEIARL 564

```

Search completed: May 8, 2002, 07:20:16  
Job time: 144 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:20:33 ; Search time 12.86 seconds  
(without alignments)  
19.958 Million cell updates/sec

Title: US-09-765-086-207  
Perfect score: 30  
Sequence: 1 SMSIARL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	26	86.7	282	1 MIND_CHLVU	P56346 chlorella v
2	26	86.7	286	1 MIND_MESVI	Q9mum5 mesostigma
3	26	86.7	748	1 KHL1_HUMAN	Q9nr64 homo sapien
4	26	86.7	751	1 KHL1_MOUSE	Q9ji74 mus musculus
5	26	86.7	831	1 YPDD_ECOLI	P77439 escherichia
6	25	83.3	117	1 RSL3_ECOLI	P02369 escherichia
7	25	83.3	126	1 RSL3_PORPU	P51295 porphyra pu
8	25	83.3	214	1 YPUS_RHOCA	P18661 rhodobacter
9	24	80.0	156	1 RS7_SYNPE	P23361 synecococc
10	24	80.0	253	1 ADH_DROPI	P44996 haemophilus
11	24	80.0	268	1 YIAJ_HAEIN	P37499 bacillus su
12	24	80.0	278	1 YYPE_BACSU	P42786 neisseria g
13	24	80.0	310	1 PIP_NEIGO	Q9juv1 neisseria m
14	24	80.0	310	1 PIP_NEIMB	Q9jzr6 methanobact
15	24	80.0	331	1 YB94_METTH	O27262 methanobact
16	24	80.0	390	1 GSPE_XANCP	P31744 xanthomonas
17	24	80.0	488	1 P331_ARATH	Q9lnu4 arabidopsis
18	24	80.0	488	1 P331_ARATH	Q9lnu4 arabidopsis
19	24	80.0	488	1 P331_ARATH	Q9lnu4 arabidopsis
20	24	80.0	516	1 CALR_RAT	P32214 rattus norv
21	24	80.0	533	1 PYRG_BORBU	O51522 borrelia bu
22	24	80.0	789	1 VIB4_AGR9	P05353 agrobacteri
23	24	80.0	859	1 YQD3_CAEEL	Q09263 caenorhabdi
24	24	80.0	1197	1 DPOM_PODAN	Q01529 podospora a
25	24	80.0	1315	1 CHAO_DROME	P12024 drosophila
26	24	80.0	1752	1 RPBI_SCHPO	P36594 schizosacch
27	23	76.7	91	1 KPYK_LEIBR	Q04668 leishmania
28	23	76.7	187	1 DPA6_BPR69	Q01743 bacterioph
29	23	76.7	215	1 SC1_OCTDO	P27009 octopus dof
30	23	76.7	215	1 SC2_OCTVU	P27014 octopus vul
31	23	76.7	215	1 SC3_OCTVU	Q25626 octopus vul
32	23	76.7	219	1 YLP4_ZYMMO	O66114 zymomonas m
33	23	76.7	220	1 KPYK_THELI	Q56301 thermococcu

34 23 76.7 313 1 NUDC\_MYCTU  
35 23 76.7 320 1 LDH\_BACSU  
36 23 76.7 321 1 YXTL\_CAEEL  
37 23 76.7 354 1 NULM\_PECMA  
38 23 76.7 392 1 YG59\_SYNY3  
39 23 76.7 392 1 YL78\_ARCFU  
40 23 76.7 452 1 HEMF\_RHOSH  
41 23 76.7 457 1 THIC\_AQUAE  
42 23 76.7 467 1 SELA\_EUBAC  
43 23 76.7 475 1 MOTD\_RHIME  
44 23 76.7 484 1 FKHI\_YEAST  
45 23 76.7 499 1 KPY1\_TRYBB

## ALIGNMENTS

RESULT 1  
MIND\_CHLVU  
ID MIND\_CHLVU STANDARD; PRT; 282 AA.  
AC P56346;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PUTATIVE SEPTUM SITE-DETERMINING PROTEIN MIND.  
GN MIND.  
OS Chlorella vulgaris.  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;  
OC Chlorellaceae; Chlorella.  
OX NCBI\_TaxID=3077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IAM C-27 / TAMIYA;  
RX MEDLINE=97303241; PubMed=9159184;  
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,  
Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A.,  
Inamura A., Yoshinaga K., Sugitara M.;  
RT "Complete nucleotide sequence of the chloroplast genome from the  
green alga Chlorella vulgaris: the existence of genes possibly  
involved in chloroplast division.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
CC -!- FUNCTION: ATPASE REQUIRED FOR THE CORRECT PLACEMENT OF THE  
DIVISION SITE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE PARA FAMILY. MIND SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed, usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB001684; BAA57951.1;  
DR InterPro; IPR000707; Para.  
DR Pfam; PF00991; Para; 1  
DR Cell division; Septation; ATP-binding; Chloroplast.  
KW NP\_BIND 24 31  
FT BIND 24 31 ATP (POTENTIAL).  
SQ SEQUENCE 282 AA; 31013 MW; 94010DD45AE4AEC7 CRC64;

Query Match 86.7%; Score 26; DB 1; Length 282;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSIARL 7  
|||||  
Db 39 MSIARL 44

RESULT 2  
MIND\_MESVI

```
ID MIND_MESVI STANDARD; PRT; 286 AA.
AC Q9MUM5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE SEPTUM SITE-DETERMINING PROTEIN MIND.
GN MIND.
OS Mesostigma viride.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Mesostigmatophyceae;
OC Mesostigmatales; Mesostigmataceae; Mesostigma.
OX NCBI_TaxID=41882;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIES-296;
RX MEDLINE=20150907; PubMed=10688199;
RA Lemieux C., Otis C., Turmel M.;
RT "Ancestral chloroplast genome in Mesostigma viride reveals an early
branch of green plant evolution.";
RL Nature 403:649-652(2000).
CC -!- FUNCTION: ATPASE REQUIRED FOR THE CORRECT PLACEMENT OF THE
DIVISION SITE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PARA FAMILY. MIND SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF166114; AAF43873.1; -.
DR InterPro; IPR000707; Para.
DR Pfam; PF00991; Para; 1.
KW Cell division; Septation; ATP-binding; Chloroplast.
FT NP_BIND 25 32 ATP (POTENTIAL).
SQ SEQUENCE 286 AA; 31627 MW; 25D8B8FB2258E3F9 CRC64;

Query Match 86.7%; Score 26; DB 1; Length 286;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSIALR 7
Db 40 MSIALR 45

RESULT 3
KHL1_HUMAN STANDARD; PRT; 748 AA.
AC Q9NR64; Q9NR65; Q9P238; Q9HAX4;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE KELCH-LIKE PROTEIN 1.
GN KHL1 OR KIAA1490.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347694; PubMed=10888605;
RA Koob M.D., Nemes J.P., Benzow K.A.;
RT "The SCAR transcript is an antisense RNA to a brain-specific
transcript encoding a novel actin-binding protein (KHL1).";
RL Hum. Mol. Genet. 9:1543-1551(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [3]
RP SEQUENCE OF 179-409 FROM N.A.
RA Kay M.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
THE BRAIN CELLS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF252283; AAF81719.1; -.
DR EMBL; AF252279; AAF81716.1; -.
DR EMBL; AB040923; BAA96014.1; ALT_INIT.
DR EMBL; AL353738; CAC16128.1; -.
DR MIM; 605332; -.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF01344; Kelch; 6.
DR PRINTS; PR00501; KELCHREPEAT.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS50097; BTB; 1.
KW Cytoskeleton; Actin-binding; Repeat.
FT DOMAIN 43 88 SER-RICH.
FT DOMAIN 212 279 BTB.
FT REPEAT 460 506 KELCH 1.
FT REPEAT 507 553 KELCH 2.
FT REPEAT 555 600 KELCH 3.
FT REPEAT 601 647 KELCH 4.
FT REPEAT 649 700 KELCH 5.
FT REPEAT 701 747 KELCH 6.
SQ SEQUENCE 748 AA; 82680 MW; C11C43D8282F9FF9 CRC64;

Query Match 86.7%; Score 26; DB 1; Length 748;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIAR 6
Db 585 SMSIAR 590

RESULT 4
KHL1_MOUSE STANDARD; PRT; 751 AA.
AC Q9J174;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE KELCH-LIKE PROTEIN 1.
GN KHL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347694; PubMed=10888605;
```

RA Koob M.D., Nemes J.P., Benzow K.A.;  
 RT "The SCAB transcript is an antisense RNA to a brain-specific  
 RL transcript encoding a novel actin-binding protein (KLHL1).";  
 CC Hum. Mol. Genet. 9:1543-1551(2000)  
 CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF  
 CC THE BRAIN CELLS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.  
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF252281; AAF81717.1; -  
 DR InterPro: IPR000210; BTB\_POZ.  
 DR InterPro: IPR001798; Kelch.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF01344; Kelch; 6.  
 DR PRINTS; PR00501; KELCHREPEAT.  
 DR SMART; SM00225; BTB; 1.  
 DR PROSITE; PS50097; BTB; 1.  
 KW Cytoskeleton; Actin-binding; Repeat.  
 FT DOMAIN 45 90  
 FT REPEAT 215 282 SER-RICH.  
 FT REPEAT 463 509  
 FT REPEAT 510 556 KELCH 1.  
 FT REPEAT 558 603 KELCH 2.  
 FT REPEAT 604 650 KELCH 3.  
 FT REPEAT 652 703 KELCH 4.  
 FT REPEAT 704 750 KELCH 5.  
 FT REPEAT 751 AA; 82932 MW; 6E846190CAC6B4C3 CRC64;  
 SQ SEQUENCE 751 AA; 82932 MW; 6E846190CAC6B4C3 CRC64;  
 -----  
 Query Match 86.7%; Score 26; DB 1; Length 751;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 1 SMSIAR 6  
 |||||  
 Db 588 SMSIAR 593  
 -----  
 RESULT 5  
 YPDD\_ECOLI STANDARD; PRT; 831 AA.  
 AC P77439; (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PUTATIVE PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE YPDD  
 DE (EC 2.7.3.9) (PHOSPHOTRANSFERASE SYSTEM, ENZYME I).  
 YPDD OR B2383.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 -----  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MGI1655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97349980; PubMed-9205837;  
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Nakano K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horiuchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 CC -1- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + PROTEIN HISTIDINE =  
 CC PYRUVATE + PROTEIN N-PHOSPHOHISTIDINE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HPR DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE000326; AAC75442.1; -  
 DR EMBL; D90868; BAA16253.1; -  
 DR EcoGene; EG14151; Ypdd.  
 DR InterPro: IPR000121; PEP-utilizers.  
 DR InterPro: IPR002178; PTS\_EIIA\_2.  
 DR Pfam; PF00391; PEP-utilizers; 1.  
 DR ProDom; PD000940; PEP-utilizers; 1.  
 DR ProDom; PD001689; PTS\_EIIA\_2; 1.  
 DR PROSITE; PS00370; PEP\_ENZYMES\_PHOS\_SITE; FALSE\_NEG.  
 DR PROSITE; PS00742; PEP\_ENZYMES\_2; 1.  
 DR PROSITE; PS00369; PTS\_HPR\_HIS; FALSE\_NEG.  
 DR PROSITE; PS00589; PTS\_HPR\_SER; FALSE\_NEG.  
 KW Hypothetical protein; Phosphotransferase system; Transferase; Kinase;  
 KW Sugar transport; Phosphorylation; Complete proteome.  
 FT DOMAIN 1 88  
 FT HPR.  
 FT DOMAIN 266 613 ENZYME 1 DOMAIN.  
 FT DOMAIN 730 831 EIIA DOMAIN.  
 FT MOD\_RES 15 15 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 298 298 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 831 AA; 92129 MW; B9F3E3B6D4EAB597 CRC64;  
 -----  
 Query Match 86.7%; Score 26; DB 1; Length 831;  
 Best Local Similarity 85.7%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 QY 1 SMSIARL 7  
 |:|||||  
 Db 757 SISIARL 763  
 -----  
 RESULT 6  
 RS13\_ECOLI STANDARD; PRT; 117 AA.  
 ID RS13\_ECOLI  
 AC P02369;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE 30S RIBOSOMAL PROTEIN S13.  
 GN RPSM OR B3298.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]

```
RP SEQUENCE FROM N.A.
RX MEDLINE-85242076; PubMed-2989779;
RA Bedwell D.M., Davis G.R., Gosink M., Post L.E., Nomura M., Kestler H.,
RT Zengel J.M., Lindahl L.;
RN "Nucleotide sequence of the alpha ribosomal protein operon of
RL Escherichia coli.";
RN Nucleic Acids Res. 13:3891-3903(1985).
RP SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RT Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RP SEQUENCE.
RP STRAIN-K;
RX MEDLINE-77248097; PubMed-330375;
RA Lindemann H., Wittmann-Liebold B.;
RT "Primary structure of protein S13 from the small subunit of
RL Escherichia coli ribosomes.";
RL Hoppe-Seyler's Z. Physiol. Chem. 358:843-863(1977).
RP SEQUENCE OF 1-35 FROM N.A.
RX MEDLINE-80182128; PubMed-6154696;
RA Post L.E., Arfsten A.E., Davis G.R., Nomura M.;
RT "DNA sequence of the promoter region for the alpha ribosomal protein
operon in Escherichia coli.";
RL J. Biol. Chem. 255:4653-4659(1980).
RP SEQUENCE OF 1-35 FROM N.A.
RX MEDLINE-82025589; PubMed-6793240;
RA Miura A., Krueger J.H., Itoh S., de Boer H.A., Nomura M.;
RT "Growth-rate-dependent regulation of ribosome synthesis in E. coli:
expression of the lacZ and galK genes fused to ribosomal promoters.";
RL Cell 25:773-782(1981).
RP MASS SPECTROMETRY.
RX MEDLINE-99196679; PubMed-10094780;
RA Arnold R.J., Keilly J.P.;
RT "Observation of Escherichia coli ribosomal proteins and their
posttranslational modifications by mass spectrometry.";
RL Anal. Biochem. 269:105-112(1999).
CC -!- FUNCTION: INVOLVED IN THE BINDING OF FMET-TRNA AND, HENCE, IN THE
INITIATION OF TRANSLATION.
CC -!- MASS SPECTROMETRY: MW-12968.1; METHOD=MALDI.
CC -!- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M12432; AAA83903.1; -
DR EMBL; U18997; AAA58093.1; -
DR EMBL; AE000407; AAC76323.1; -
DR EMBL; X02543; CAA26392.1; -
DR EMBL; M10213; AAA72457.1; -
DR PIR; A23807; R3EC13.
DR EcGene; Egl0912; rpsM.
DR InterPro; IPR001892; Ribosomal_S13.
DR Pfam; PF00416; Ribosomal_S13; 1.
DR ProDom; PD001363; Ribosomal_S13; 1.
DR PROSITE; PS00646; RIBOSOMAL_S13; 1.
KW Ribosomal protein; Complete proteome.
FT INIT_MET 0
SQ SEQUENCE 117 AA; 12968 MW; 3277C328EBD0D3D9 CRC64;

SEQUENCE FROM N.A.
Query Match 83.3%; Score 25; DB 1; Length 117;
Best Local Similarity 85.7%; Pred. No. 9.9;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7
DB 73 SMSIKRL 79

RESULT 7
RR13_PORPU
ID RR13_PORPU STANDARD; PRT; 126 AA.
AC P51295;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S13.
GN RPS13.
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
OX NCBI_TaxID-2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AVONPORT;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -!- FUNCTION: INVOLVED IN THE BINDING OF FMET-TRNA AND, HENCE, IN THE
INITIATION OF TRANSLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U38804; AAC08181.1; -
DR Mendel; 10345; PORPU; rps13; 1.
DR InterPro; IPR001892; Ribosomal_S13.
DR Pfam; PF00416; Ribosomal_S13; 1.
DR ProDom; PD001363; Ribosomal_S13; 1.
DR PROSITE; PS00646; RIBOSOMAL_S13; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 126 AA; 14538 MW; D7C26927606FE12B CRC64;

Query Match 83.3%; Score 25; DB 1; Length 126;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7
DB 75 SMSIKRL 81

RESULT 8
YPU5_RHOCA
ID YPU5_RHOCA STANDARD; PRT; 214 AA.
AC P26161;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE HYPOTHETICAL 23.7 KDA PROTEIN IN PUHA 5'REGION (ORF214) (PROTEIN
F3981).
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
```

```
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RA Burke D.H., Alberti M., Armstrong G.A., Hearst J.E.;
RL Submitted (Nov-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-14 FROM N.A.
RX MEDLINE=84259352; PubMed=6744416;
RA Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
RT "Nucleotide and deduced polypeptide sequences of the photosynthetic
RT reaction-center, B870 antenna, and flanking polypeptides from R.
RT capsulata.";
RL Cell 37:949-957(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11165; CAA77519.1; -
DR EMBL; K01183; -; NOT_ANNOTATED_CDS.
DR PIR; S17807; S17807.
DR PIR; C28988; C28988.
KW Photosynthesis; Hypothetical protein.
SQ SEQUENCE 214 AA; 23658 MW; BC4B773AF6631FA3 CRC64;

Query Match 83.3%; Score 25; DB 1; Length 214;
Best Local Similarity 71.4%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
DB 129 SMSLAKL 135
||||:|

RESULT 9
RS7_SYNP6 STANDARD; PRT; 156 AA.
AC P18661;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S7.
GN RPSG.
OS Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1139;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89281486; PubMed=2499762;
RA Meng B.Y., Shinozaki K., Sugliura M.;
RT "Genes for the ribosomal proteins S12 and S7 and elongation factors
RT EF-G and EF-Tu of the cyanobacterium, Anacystis nidulans: structural
RT homology between 16S rRNA and S7 mRNA.";
RL Mol. Gen. Genet. 216:25-30(1989).
CC -!- FUNCTION: PROTEIN S7 BINDS SPECIFICALLY TO PART OF THE 3' END OF
CC 16S RIBOSOMAL RNA.
CC -!- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17442; CAA35494.1; -
```

```
DR PIR; S04428; S04428.
DR HSSP; P22744; 1HUS.
DR InterPro; IPR000235; Ribosomal_S7.
DR Pfam; PF00177; Ribosomal_S7; 1.
DR PRODOM; PD000817; Ribosomal_S7; 1.
DR PROSITE; PS00052; RIBOSOMAL_S7; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 156 AA; 17734 MW; 240FE99581CADCCC CRC64;

Query Match 80.0%; Score 24; DB 1; Length 156;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7
DB 24 SMVVARL 30
|||||

RESULT 10
ADH_DROPI STANDARD; PRT; 253 AA.
AC P23361;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALCOHOL DEHYDROGENASE (EC 1.1.1.1).
GN ADH.
OS Drosophila plecticornis (Fruit fly) (Idiomyla plecticornis).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7235;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91163323; PubMed=2002765;
RA Rowan R.G., Hunt J.A.;
RT "Rates of DNA change and phylogeny from the DNA sequences of the
RT alcohol dehydrogenase gene for five closely related species of
RT Hawaiian Drosophila.";
RL Mol. Biol. Evol. 8:49-70(1991).
CC -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) -> ALDEHYDE OR KETONE + NADH.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63392; AAA28353.1; -
DR PIR; B23724; B23724.
DR FlyBase; FBan0012651; Dpic\adh.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR003030; ADH_short_C.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00663; adh_short_C; 1.
DR PRINTS; PRO0080; SDRFAMILY.
DR PRINTS; PRO1167; INSADHFAMILY.
DR PRINTS; PRO1168; ALCDHGRNASE.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NAD.
FT INIT_MET 0 0 BY SIMILARITY.
FT NP_BIND 9 32 NAD (BY SIMILARITY).
FT ACT_SITE 150 150 BY SIMILARITY.
SQ SEQUENCE 253 AA; 27437 MW; 12CDAAB491E3505B CRC64;
```

Query Match 80.0%; Score 24; DB 1; Length 253;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
DB 161 TMSIAKL 167

RESULT 11  
YIAJ\_HAEIN STANDARD; PRT; 268 AA.  
AC YIAJ\_HAEIN  
AC P44996;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR HII032.  
GN HII032.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -!- SIMILARITY: BELONGS TO THE ICRL FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS. STRONG, TO E.COLI YIAJ.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U32784; AAC22692.1; -;  
DR TIGR; HI1032; -;  
DR InterPro; IPR000285; HTH\_ICLR.  
DR Pfam; PF01614; ICLR; 1  
DR ProDom; PD002768; HTH\_ICLR; 1.  
DR SMART; SM00346; HTH\_ICLR; 1.  
DR PROSITE; PS01051; HTH\_ICLR\_FAMILY; 1.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
FT DNA\_BIND 37 56 H-T-H MOTIF (POTENTIAL).  
SQ SEQUENCE 268 AA; 30580 MW; AAAF3080ED39EA14 CRC64;

Query Match 80.0%; Score 24; DB 1; Length 268;  
Best Local Similarity 85.7%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
DB 234 SMSIVRL 240

RESULT 12  
YIBE\_BACSU STANDARD; PRT; 278 AA.  
AC P37499;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR YIBE.  
GN YIBE.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
CC -!- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D26185; BAA05198.1; -;  
DR EMBL; Z99124; CAB16104.1; -;  
DR Subtilist; BG10026; YIBE.  
DR InterPro; IPR000847; HTH\_LysR.  
DR Pfam; PF00126; HTH\_L; 1.  
DR PRINTS; PR00039; HTHLYSR.  
DR PROSITE; PS00044; HTH\_LYSR\_FAMILY; 1.  
KW Hypothetical protein; Transcription regulation; DNA-binding;  
FT DNA\_BIND 4 23 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 278 AA; 31406 MW; 0F7750102F0FC07 CRC64;

Query Match 80.0%; Score 24; DB 1; Length 278;  
Best Local Similarity 85.7%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
DB 19 SRSIARL 25

RESULT 13  
PIP\_NEIGO STANDARD; PRT; 310 AA.  
AC P42786;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROLINE IMINOPEPTIDASE (EC 3.4.11.5) (PIP) (PROLYL AMINOPEPTIDASE)  
DE (PAP).  
GN PIP.  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=485;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
RC STRAIN=MS11 / MS01-1X;  
RX MEDLINE=95020651; PubMed=7934933;  
RA Albertson N.H., Koomey M.;  
RT "Molecular cloning and characterization of a proline iminopeptidase  
RT gene from Neisseria gonorrhoeae.";  
RL Mol. Microbiol. 9:1203-1211(1993).



```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002444; AAF41334.1; -.
DR TIGR; NMB0927; -.
DR InterPro; IPR003089; AB_hydrolase.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR000639; Epox_hydrase.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR002410; Pro_aminoptase.
DR Pfam; PF00561; abhydrolase; 1.
DR PRINTS; PRO0111; ABHYDROLASE.
DR PRINTS; PRO0412; EPOXHYDRLASE.
DR PRINTS; PRO0793; PROAMINOPTASE.
KW Hydrolase; Aminopeptidase; Complete proteome.
FT ACT_SITE 107 107 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 260 260 BY SIMILARITY.
FT ACT_SITE 287 287 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 310 AA; 34956 MW; 9F883415B1C7B4A2 CRC64;

Query Match      80.0%; Score 24; DB 1; Length 310;
Best Local Similarity 71.4%; Pred. No. 53;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 217 SLAIARL 223
```

Search completed: May 8, 2002, 07:20:34  
Job time: 162 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:19:31 ; Search time 22.68 seconds  
(without alignments)  
23.511 Million cell updates/sec

Title: US-09-765-086-207  
Perfect score: 30  
Sequence: 1 SMSIARL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 7617452 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	27	90.0	301	2	D81745	probable phosphati
2	27	90.0	301	2	E71482	phosphatidylserine
3	27	90.0	329	2	G86472	probable hyosciami
4	27	90.0	369	2	T05013	hypothetical prote
5	26	86.7	282	2	T07303	cell division inhi
6	26	86.7	305	2	G72588	probable dTDP-4-de
7	26	86.7	307	2	T03502	conserved hypothet
8	26	86.7	422	2	I50516	retinoid x recepto
9	26	86.7	831	2	A85881	hypothetical prote
10	26	86.7	831	2	D65012	hypothetical prote
11	26	86.7	849	2	S61631	probable membrane
12	26	86.7	1216	2	H85023	hypothetical prote
13	25	83.3	98	2	G85926	hypothetical prote
14	25	83.3	118	1	R3EC13	ribosomal protein
15	25	83.3	118	2	G85994	30S ribosomal subu
16	25	83.3	126	2	S73216	ribosomal protein
17	25	83.3	213	2	D86170	hypothetical prote
18	25	83.3	214	2	S17807	hypothetical prote
19	25	83.3	255	2	T38115	probable ATP-depen
20	25	83.3	283	2	C83338	hypothetical prote
21	25	83.3	294	2	S76046	hypothetical prote
22	25	83.3	539	2	T47405	hypothetical prote
23	25	83.3	613	2	T00077	gag-like protein -
24	24	80.0	98	2	T42264	hypothetical prote
25	24	80.0	107	2	G83348	hypothetical prote
26	24	80.0	156	2	S04428	ribosomal protein
27	24	80.0	223	1	B23724	alcohol dehydrogen
28	24	80.0	266	2	D71945	hypothetical prote
29	24	80.0	268	2	D64165	hypothetical prote

30 24 80.0 270 2 E64561 DNA processing cha  
31 24 80.0 278 2 S65992 transcription regu  
32 24 80.0 299 2 T32333 hypothetical prote  
33 24 80.0 301 2 F82446 transcription regu  
34 24 80.0 310 2 S39592 prolyl aminopeptid  
35 24 80.0 310 2 F81878 probable prolyl am  
36 24 80.0 310 2 B81141 proline iminopepti  
37 24 80.0 313 2 F72575 hypothetical prote  
38 24 80.0 318 2 A83708 hypothetical prote  
39 24 80.0 331 1 C69026 acetylpolymamine am  
40 24 80.0 336 2 T01839 hypothetical prote  
41 24 80.0 390 1 S17938 xpsf protein - xan  
42 24 80.0 390 2 T12057 xpsf protein - xan  
43 24 80.0 478 2 A37430 calcitonin recepto  
44 24 80.0 479 2 S3746 calcitonin recepto  
45 24 80.0 488 2 T02207 protein 21D7 - com

ALIGNMENTS

RESULT 1

D81745

probable phosphatidylserine decarboxylase (EC 4.1.1.65) precursor TC0072 [similarity]  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 18-Aug-2000  
C:Accession: D81745  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
, C.; Dodson, R.; Gwinn, M.; Neilson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: D81745  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-301 <TET>  
A:Cross-references: GB:AE002275; GB:AE002160; NID:g7190108; PIDN:AAF38954.1; PID:g719  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0072  
C:Superfamily: Escherichia coli phosphatidylserine decarboxylase  
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 90.0%; Score 27; DB 2; Length 301;  
Best Local Similarity 85.7%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARL 7

Db 158 SMAIARL 164

RESULT 2

E71482

phosphatidylserine decarboxylase (EC 4.1.1.65) precursor - Chlamydia trachomatis (ser  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 28-Jul-2000  
C:Accession: E71482  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: E71482  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-301 <ARN>  
A:Cross-references: GB:AE001340; GB:AE001273; NID:g3329147; PIDN:AAC68294.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: psdD  
C:Superfamily: Escherichia coli phosphatidylserine decarboxylase  
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 90.0%; Score 27; DB 2; Length 301;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 ||:||||  
 Db 158 SMAIARL 164

RESULT 3  
 G86472  
 probable hycosamine 6-dioxygenase hydroxylase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: G86472  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: G86472  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-329 <STO>  
 A:Cross-references: GB:AE005172; NID:g11386306; PIDN:AAG35097.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: 1-aminocyclopropane-1-carboxylate oxidase

Query Match 90.0%; Score 27; DB 2; Length 329;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 ||:||||  
 Db 150 SMAIARL 156

RESULT 4  
 T05013  
 hypothetical protein T19P19.150 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999  
 C:Accession: T05013  
 R:Bevan, M.; Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Hoheisel, J.; Mewes, H.W.;  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: Z15394  
 A:Accession: T05013  
 A:Molecule type: DNA  
 A:Residues: 1-369 <BEV>  
 A:Cross-references: EMBL:AL022605  
 A:Experimental source: cultivar Columbia; BAC clone T19P19  
 C:Genetics:  
 A:Map position: 4  
 A:Note: T19P19.150  
 C:Superfamily: coffeoyl-CoA 3-O-methyltransferase

Query Match 90.0%; Score 27; DB 2; Length 369;  
 Best Local Similarity 71.4%; Pred. No. 27;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 ||:||||

Db 162 SMSVARI 168

RESULT 5  
 T07303  
 cell division inhibitor - Chlorella vulgaris chloroplast  
 C:Species: chloroplast Chlorella vulgaris  
 C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000  
 C:Accession: T07303  
 R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Na  
 Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997  
 A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga C  
 A:Reference number: Z15985; MUID:97303241  
 A:Accession: T07303  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-282 <WAK>  
 A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAA57951.1; PID:g2224467  
 C:Genetics:  
 A:Gene: mind  
 A:Genome: chloroplast  
 C:Superfamily: cell division inhibitor mind  
 C:Keywords: Chloroplast

Query Match 86.7%; Score 26; DB 2; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSIARL 7  
 ||:||||  
 Db 39 MSIARL 44

RESULT 6  
 G72588  
 probable dTDP-4-dehydrorhamnose reductase APE1179 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: G72588  
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
 A:Reference number: A72450; MUID:99310339  
 A:Accession: G72588  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-305 <KAW>  
 A:Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80165.1; PID:g5104851  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE1179  
 C:Superfamily: dTDP-dihydrostreptose synthase

Query Match 86.7%; Score 26; DB 2; Length 305;  
 Best Local Similarity 71.4%; Pred. No. 41;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 ||:||||  
 Db 99 SMTVARL 105

RESULT 7  
 T03502  
 conserved hypothetical protein - Rhodobacter capsulatus  
 C:Species: Rhodobacter capsulatus  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999  
 C:Accession: T03502  
 R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997

A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003  
A:Reference number: 214955; MUID:97404404  
A:Accession: T03502  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-307 <VLC>  
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16155.1; PID:g3128303  
C:Genetics:  
A:Map position: 1  
C:Superfamily: conserved hypothetical protein HI0963

Query Match 86.7%; Score 26; DB 2; Length 307;  
Best Local Similarity 71.4%; Pred. No. 41;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARL 7  
Db 205 NMSVARL 211  
:|||||

RESULT 8  
I50516  
retinoid X receptor delta - zebra fish  
C:Species: Brachydanio rerio (zebra fish)  
C:Date: 13-Mar-1997 #sequence\_revision 13-Mar-1997 #text\_change 20-Sep-1999  
C:Accession: I50516  
R:Jones, B.B.; Ohno, C.K.; Allenby, G.; Boffa, M.B.; Levin, A.A.; Grippo, J.F.; Petkovic  
Mol. Cell. Biol. 15, 5226-5234, 1995  
A:Title: New retinoid X receptor subtypes in zebra fish (Danio rerio) differentially mod  
A:Reference number: A57301; MUID:96009547  
A:Accession: I50516  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-422 <JON>  
A:Cross-references: EMBL:U29941; NID:g1046296; PIDN:AAC59721.1; PID:g1046297  
C:Genetics:  
A:Gene: RXR  
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology  
C:Keywords: zinc finger  
F:88-341/Domain: erba transforming protein homology <ERBA>

Query Match 86.7%; Score 26; DB 2; Length 422;  
Best Local Similarity 71.4%; Pred. No. 58;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARL 7  
Db 53 SMSVSRL 59  
:|||||

RESULT 9  
A85881  
hypothetical protein Z3648 [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: A85881  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A85881  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-831 <STO>  
A:Cross-references: GB:AE005174; NID:g12516757; PIDN:AAG57509.1; GSPDB:GN00145; UWGP:Z36  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z3648

Query Match 86.7%; Score 26; DB 2; Length 831;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARL 7  
Db 757 SISIARL 763  
:|||||

RESULT 10  
D65012  
hypothetical protein b2383 - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C:Accession: D65012  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: D65012  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-831 <BLAT>  
A:Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AAC75442.1; PID:g17887  
A:Experimental source: strain K-12, substrain M61655  
C:Superfamily: phosphotransferase system enzyme I homology  
F:118-670/Domain: phosphotransferase system enzyme I homology <PTI>

Query Match 86.7%; Score 26; DB 2; Length 831;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARL 7  
Db 757 SISIARL 763  
:|||||

RESULT 11  
S61631  
probable membrane protein YLR057w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein L2153  
C:Species: Saccharomyces cerevisiae  
C:Date: 09-Mar-1996 #sequence\_revision 12-Apr-1996 #text\_change 21-Nov-1997  
C:Accession: S61631; S64885  
R:Urrestazu, L.A.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: S61618  
A:Accession: S61631  
A:Molecule type: DNA  
A:Residues: 1-849 <URR>  
A:Cross-references: EMBL:X94607; NID:g1181264; PID:e217771; PID:g1181278  
R:Andre, B.; Urrestazu, L.A.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64872  
A:Accession: S64885  
A:Molecule type: DNA  
A:Residues: 1-849 <AND>  
A:Cross-references: EMBL:Z73229; NID:g1360399; PID:e245524; PID:g1360400; MIPS:YLR057  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 12R  
C:Keywords: transmembrane protein  
F:16-32/Domain: transmembrane #status predicted <TM>

Query Match 86.7%; Score 26; DB 2; Length 849;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 MSIARL 7  
:|||||

Db 1 MSIALR 6

RESULT 12

H85023

hypothetical protein AT4g01860 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001

C:Accession: H85023

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: H85023

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1216 <STO>

A:Cross-references: GB:NC\_001268; NID:g7268570; PIDN:CAB80679.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g01860

A:Map position: 4

Query Match 86.7%; Score 26; DB 2; Length 1216;

Best Local Similarity 71.4%; Pred. No. 1.8e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIALR 7

Db 601 SLSVARL 607

RESULT 13

G85926

hypothetical protein Z4077 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: G85926

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85926

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-98 <STO>

A:Cross-references: GB:AE005174; NID:g12517228; PIDN:AAG57875.1; GSPDB:GN00145; UWGP:Z40

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z4077

Query Match 83.3%; Score 25; DB 2; Length 98;

Best Local Similarity 83.3%; Pred. No. 23;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIALR 6

Db 12 SMSVAR 17

RESULT 14

R3EC13

ribosomal protein S13 [validated] - Escherichia coli

C:Species: Escherichia coli

C>Date: 24-Apr-1984 #sequence\_revision 30-Jun-1991 #text\_change 28-Jan-2000

R:Bedwell, D.; Davis, G.; Gosink, M.; Post, L.; Nomura, M.; Kestler, H.; Zengel, J.M.; I Nucleic Acids Res. 13, 3891-3903, 1985

A:Title: Nucleotide sequence of the alpha ribosomal protein operon of Escherichia coli.

A:Reference number: A23807; MUID:85242076

A:Accession: A23807

A:Gene: rpsM

A:Molecule type: DNA

A:Residues: 1-118 <BED>

A:Cross-references: GB:X02543; NID:g42795; PIDN:CAA26392.1; PID:g581217

A>Note: the authors translated the initiation codon GTG for residue 1 as Val

R:Lindemann, H.; Wittmann-Liebold, B.

Hoppe-Seyler's Z. Physiol. Chem. 358, 843-863, 1977

A:Title: Primary structure of protein S13 from the small subunit of Escherichia coli

A:Reference number: A02731; MUID:77248097

A:Accession: A02731

A:Molecule type: protein

A:Residues: 2-118 <LIN>

A:Experimental source: strain K

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; .A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: E65122

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-118 <BLAT>

A:Cross-references: GB:AE000407; GB:U00096; NID:g2367211; PIDN:AAC76323.1; PID:g17896

A:Experimental source: strain K-12, substrain MG1655

R:Arnold, R.J.; Reilly, J.P.

Anal. Biochem. 269, 105-112, 1999

A:Title: Observation of Escherichia coli ribosomal proteins and their posttranslation

A:Reference number: A59071; MUID:99196679

A:Contents: annotation; mass spectrographic analysis

A>Note: mass spectrographic analysis of post-translational modifications; any acid la

C:Genetics:

A:Gene: rpsM

A:Map position: 73 min

A:Start codon: GTG

C:Complex: the ribosome is composed of the large (50S) and small (30S) subunit; the 1

S rRNA and 22 distinct proteins

C:Complex: small subunit ribosomal proteins: S1 (PIR:R3EC1), S2 (PIR:R3EC2), S3 (PIR: R3EC10), S11 (PIR:R3EC11), S12 (PIR:R3EC12), S13 (PIR:R3EC13), S14 (PIR:R3EC14), S15

IR:R3EC21), S22 (PIR:C64901) [validated; MUID:99196679]

C:Function:

A:Pathway: protein biosynthesis

C:Superfamily: Escherichia coli ribosomal protein S13

C:Keywords: protein biosynthesis; ribosome

F;2-118/product: ribosomal protein S13 #status experimental <MAT>

Query Match 83.3%; Score 25; DB 1; Length 118;

Best Local Similarity 85.7%; Pred. No. 28;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIALR 7

Db 74 SMSIKRL 80

RESULT 15

G85994

30S ribosomal subunit protein S13 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: G85994

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85994

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-118 <STO>

A:Cross-references: GB:AE005174; NID:g12517927; PIDN:AAG58419.1; GSPDB:GN00145; UWGP: A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: rpsM

C:Superfamily: Escherichia coli ribosomal protein S13

Query Match 83.3%; Score 25; DB 2; Length 118;  
Best Local Similarity 85.7%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SMSIARL 7  
    |||||  
Db 74 SMSIKRL 80

Search completed: May 8, 2002, 07:19:33  
Job time: 101 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:17:52 ; Search time 35.89 seconds  
(without alignments)  
57.058 Million cell updates/sec

Title: US-09-765-086-200

Perfect score: 62

Sequence: 1 KLAKLAKKLAKLAK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	67.7	805	2 Q99XT3	Q99xt3 streptococc
2	41	66.1	389	5 Q9V6R9	Q9v6r9 drosophila
3	40.5	65.3	476	10 Q9C7V6	Q9c7v6 arabidopsis
4	40	64.5	222	2 Q9JQ09	Q9jq09 neisseria m
5	40	64.5	426	2 Q9X1H8	Q9x1h8 thermotoga
6	39.5	63.7	648	5 Q20191	Q20191 caenorhabdi
7	39	62.9	208	2 P95826	P95826 streptococc
8	39	62.9	436	2 Q54839	Q54839 streptococc
9	39	62.9	441	2 Q55246	Q55246 streptococc
10	39	62.9	454	2 Q55278	Q55278 streptococc
11	39	62.9	455	2 Q9PMR1	Q9pmr1 campylobact
12	39	62.9	472	2 Q55279	Q55279 streptococc
13	39	62.9	475	2 Q33631	Q33631 streptococc
14	39	62.9	532	2 Q55098	Q55098 streptococc
15	39	62.9	550	2 Q54840	Q54840 streptococc
16	39	62.9	587	2 Q55312	Q55312 streptococc
17	39	62.9	592	2 Q00720	Q00720 group g str
18	39	62.9	592	2 Q9L4N1	Q9l4n1 streptococc
19	38	61.3	276	2 Q9K5D4	Q9k5d4 campylobact

20	38	61.3	461	3 Q03785	Q03785 saccharomyc
21	38	61.3	462	5 Q09958	Q09958 caenorhabdi
22	38	61.3	1184	3 Q9Y7U0	Q9y7u0 schizosacch
23	37	59.7	92	9 Q9AZ19	Q9az19 bacterioph
24	37	59.7	105	5 Q9GSE7	Q9gse7 dictyosteli
25	37	59.7	127	1 Q9YCP7	Q9ycp7 aeropyrum p
26	37	59.7	174	2 Q48758	Q48758 listeria mo
27	37	59.7	246	5 Q93548	Q93548 caenorhabdi
28	37	59.7	282	2 Q9K6L4	Q9k6l4 bacillus ha
29	37	59.7	347	2 Q9JSP0	Q9jsp0 neisseria m
30	37	59.7	393	1 Q58728	Q58728 methanococc
31	37	59.7	412	5 Q20873	Q20873 caenorhabdi
32	37	59.7	457	2 Q99R33	Q99r33 staphylococ
33	37	59.7	537	3 Q14236	Q14236 schizosacch
34	37	59.7	684	5 Q9VXS2	Q9vxs2 drosophila
35	37	59.7	758	1 Q9UY62	Q9uy62 pyrococcus
36	37	59.7	1050	5 Q9UIQ4	Q9uiq4 caenorhabdi
37	37	59.7	1240	12 Q9DWH8	Q9dwh8 rat cytomeg
38	37	59.7	1272	5 Q9N4H7	Q9n4h7 caenorhabdi
39	37	59.7	1386	8 Q9XMS2	Q9xms2 tetrahymena
40	36	58.1	198	5 Q77005	Q77005 plasmodium
41	36	58.1	198	5 Q9TY39	Q9ty39 plasmodium
42	36	58.1	218	2 Q67003	Q67003 aquifex aeo
43	36	58.1	227	4 Q14222	Q14222 homo sapien
44	36	58.1	249	2 Q99XW9	Q99xw9 streptococc
45	36	58.1	295	5 Q9UIB6	Q9uib6 leishmania

#### ALIGNMENTS

RESULT 1

Q99XT3 ID Q99XT3 PRELIMINARY; PRT; 805 AA.  
AC Q99XT3;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PUTATIVE PYRUVATE FORMATE-LYASE 2 (EC 2.3.1.54).  
GN PFID OR SPV2049.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferreretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL; AE006625; AAK34714.1; -;  
KW Lyase; Pyruvate; Transferase; Acyltransferase; Complete proteome.  
SQ SEQUENCE 805 AA; 90464 MW; 3C41F1686D761B6E CRC64;

Query Match 67.7%; Score 42; DB 2; Length 805;  
Best Local Similarity 64.3%; Pred. No. 2e+02;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14

Db 224 RYAKLARELAKTAK 237

RESULT 2

Q9V6R9 ID Q9V6R9 PRELIMINARY; PRT; 389 AA.  
AC Q9V6R9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE CGL13333 PROTEIN.  
 GN CGL13333.  
 OS Drosophila melanogaster (Fruit fly).  
 NC Eukaryota; Metazoa; Arthropoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephygroidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cwikley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003818; AAF58353.1; -.  
 DR FlyBase; FBgn0033855; CG13333.  
 SQ SEQUENCE 389 AA; 43265 MW; 01409135653619F0 CRC64;

Query Match 66.1%; Score 41; DB 5; Length 389;  
 Best Local Similarity 69.2%; Pred. No. 1.4e+02;  
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLA 13  
 DB 339 KLAELAKRLSTLA 351  
 RESULT 3  
 Q9C7V6 PRELIMINARY; PRT; 476 AA.  
 AC Q9C7V6;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE BZIP TRANSCRIPTION FACTOR, PUTATIVE.

GN F15H21.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana";  
 RL Nature 408:816-820(2000).  
 DR EMBL; AC066689; AAC51716.1; -.  
 SQ SEQUENCE 476 AA; 55488 MW; D11027732F9C277C CRC64;  
 Query Match 65.3%; Score 40.5; DB 10; Length 476;  
 Best Local Similarity 57.1%; Pred. No. 2e+02;  
 Matches 12; Conservative 1; Mismatches 1; Indels 7; Gaps 1;  
 QY 1 KLAKLAKLA-----KLAK 14  
 DB 393 KLSKLEKLAEGTEKLLAK 413  
 RESULT 4  
 Q9JQ09 PRELIMINARY; PRT; 222 AA.  
 ID Q9JQ09;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL PROTEIN NMA1321 (BASEPLATE ASSEMBLY PROTEIN V, PUTATIVE).  
 GN NMA1321 OR NMB1111.  
 OS Neisseria meningitidis (serogroup A), and  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699, 491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead M., Warrall B.G., Barrell B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491";  
 RL Nature 404:502-506(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,



```
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoo R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citron H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AL162755; CAB84572.1; -.
DR EMBL: AE002460; AAF41502.1; -.
DR TIGR: NMB1111; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 222 AA; 23251 MW; E0AFE240175C90DF CRC64;

Query Match 64.5%; Score 40; DB 2; Length 222;
Best Local Similarity 69.2%; Pred. No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LAKLAKKLAKLAK 14
Db 3 LSKLAKTAQTAK 15

RESULT 5
Q9X1H8 PRELIMINARY; PRT; 426 AA.
AC Q9X1H8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TM1467.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=9287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL: AE001797; AAD36535.1; -.
DR TIGR: TM1467; -.
DR InterPro: IPR001179; FKBP_PPase.
DR InterPro: IPR002819; HD.
DR InterPro: IPR003607; HD.
DR Pfam: PF01966; HD; 1.
DR SMART: SM00471; HDC; 1.
DR PROSITE: PS00453; FKBP_PPase_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 426 AA; 48981 MW; 9DEAPCB7BF27E280 CRC64;

Query Match 64.5%; Score 40; DB 2; Length 426;
Best Local Similarity 72.7%; Pred. No. 2.1e+02;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLAKLAKKLAK 11
Db 260 KVALAKKLAK 270

RESULT 6
```

```
Q20191 PRELIMINARY; PRT; 648 AA.
AC Q20191; Q21661.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE F39D8.4 PROTEIN.
GN F39D8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Ainscough R.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z69791; CAA93663.1; -.
DR EMBL: Z69793; CAA93663.1; JOINED.
DR EMBL: Z69793; CAA93676.1; -.
DR EMBL: Z69791; CAA93676.1; JOINED.
DR HSSP: P28825; IIAF.
DR InterPro: IPR001506; Astacin.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003582; ShKT.
DR InterPro: IPR000130; Zn_MTpeptdse.
DR Pfam: PF01400; Astacin; 1.
DR PRINTS: PR00480; ASTACIN.
DR SMART: SM00254; ShKT; 1.
DR SMART: SM00235; ZnMC; 1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 648 AA; 73405 MW; 1661914F3CEAC71E CRC64;

Query Match 53.7%; Score 39.5; DB 5; Length 648;
Best Local Similarity 73.3%; Pred. No. 3.8e+02;
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 KLAKLAKKLAK-LAK 14
Db 376 KLAKLKGKLGKTLAK 390

RESULT 7
P95826 PRELIMINARY; PRT; 208 AA.
AC P95826;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE M PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RHD152-.
RA Brandt E.R., Good M.F.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U66005; AAB40642.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PR00015; GP0SANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Transmembrane.
FT NON_TER 1
FT NON_TER 208
SQ SEQUENCE 208 AA; 22565 MW; 79972A987324729B CRC64;

Query Match 62.9%; Score 39; DB 2; Length 208;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
```

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKL 12  
|||||:|||||  
Db 116 KLAKQAEELAKL 127

RESULT 8  
Q54839 PRELIMINARY; PRT; 436 AA.  
AC Q54839;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE M PROTEIN TYPE 52.  
GN EMM52.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus;  
OX NCBI\_TaxID=1314;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=A871/14/3;  
RX MEDLINE=93204905; PubMed=8455563;  
RA Podbielski A.;  
RT "Three different types of organization of the vir regulon in group A streptococci.";  
RL Mol. Gen. Genet. 237:287-300(1993).  
DR EMBL: X58179; CAA41168.1; -;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR003345; M\_repeat.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR Pfam: PF02370; M; 8.  
DR PRINTS: PR00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 436 AA; 48951 MW; 4C5720F98F2DAE89 CRC64;

Query Match 62.9%; Score 39; DB 2; Length 436;  
Best Local Similarity 75.0%; Pred. No. 3e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKL 12  
|||||:|||||  
Db 344 KLAKQAEELAKL 355

RESULT 9  
Q55246 PRELIMINARY; PRT; 441 AA.  
AC Q55246;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE M PROTEIN.  
GN EMM.  
OS Streptococcus sp.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus;  
OX NCBI\_TaxID=1306;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=GROUP G, 166B (ATCC 12394);  
RX MEDLINE=95229925; PubMed=7714192;  
RA Schnitzler N., Podbielski A., Baumgarten G., Mignon M., Kaufhold A.;  
RT "M or M-like protein gene polymorphisms in human group G streptococci.";  
RL J. Clin. Microbiol. 33:356-363(1995).  
DR EMBL: X60098; CAA42694.1; -;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR003345; M\_repeat.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.

DR Pfam: PF02370; M; 7.  
DR PRINTS: PR00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Actin-binding; Transmembrane.  
SQ SEQUENCE 441 AA; 49868 MW; 33DF991E365D9455 CRC64;

Query Match 62.9%; Score 39; DB 2; Length 441;  
Best Local Similarity 75.0%; Pred. No. 3.1e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKL 12  
|||||:|||||  
Db 349 KLAKQAEELAKL 360

RESULT 10  
Q55278 PRELIMINARY; PRT; 454 AA.  
AC Q55278;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE MLC36 PRECURSOR (FRAGMENT).  
OS Streptococcus sp.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus;  
OX NCBI\_TaxID=1306;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=C36;  
RA Ben Nasr A., Wistedt A., Ringdahl U., Sjobring U.;  
RL Eur. J. Biochem. 0:0-0(1994).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=C36;  
RX MEDLINE=94291620; PubMed=8020466;  
RA Ben Nasr A., Wistedt A., Ringdahl U., Sjobring U.;  
RT "Streptokinase activates plasminogen bound to human group C and G streptococci through M-like proteins.";  
RL Eur. J. Biochem. 222:267-276(1994).  
DR EMBL: Z32677; CAA83588.1; -;  
DR HSP; P03069; IGCL.  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR003345; M\_repeat.  
DR Pfam: PF02370; M; 9.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Signal.  
FT NON\_TER 1 1  
FT SIGNAL <1 29 POTENTIAL.  
FT CHAIN 30 >454 MLC36.  
FT NON\_TER 454 454  
SQ SEQUENCE 454 AA; 51416 MW; FA7D34562548282F CRC64;

Query Match 62.9%; Score 39; DB 2; Length 454;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKL 12  
|||||:|||||  
Db 389 KLAKQAEELAKL 400

RESULT 11  
Q9PMR1 PRELIMINARY; PRT; 455 AA.  
AC Q9PMR1;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE PUTATIVE FUMARATE LYASE.  
GN CJ1394.

OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCCT 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
DR EMBL; AL139078; CAB73819.1; -.  
DR InterPro; IPR000362; Fumarate\_lyase.  
DR InterPro; IPR001865; Ribosomal\_S2.  
DR Pfam; PF00206; lyase.1; 1.  
DR PRINTS; PR00149; FUMARATELYASE.  
DR PROSITE; PS00163; FUMARATE\_LYASES; 1.  
DR PROSITE; PS00962; RIBOSOMAL\_S2\_1; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 455 AA; 51371 MW; C77CA8638C84B08F CRC64;

Query Match 62.9%; Score 39; DB 2; Length 455;

Best Local Similarity 64.3%; Pred. No. 3.2e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAK 14  
DB 131 ELRAIAKALAKLAK 144  
: | | | | | | | |

RESULT 12  
ID Q55279 PRELIMINARY; PRT; 472 AA.  
AC Q55279;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE MG72 PRECURSOR.  
OS Streptococcus sp.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G72;  
RX MEDLINE=94291620; PubMed=8020466;  
RA Ben Nasr A., Wistedt A., Ringdahl U., Sjobring U.;  
RT "Streptokinase activates plasminogen bound to human group C and G  
streptococci through M-like proteins.";  
RL Eur. J. Biochem. 232:267-276(1994).  
DR EMBL; 232678; CAA83589.1; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003345; M\_repeat.  
DR Pfam; PF02370; M; 9.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Signal.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 472 MATURE MG72.  
SQ SEQUENCE 472 AA; 53968 MW; 8DE086B2F45FFC8 CRC64;

Query Match 62.9%; Score 39; DB 2; Length 472;

Best Local Similarity 75.0%; Pred. No. 3.3e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAK 12  
: | | | | | | | |

Db 407 KLAKQAEELAKL 418  
RESULT 13  
O33631  
ID O33631 PRELIMINARY; PRT; 475 AA.  
AC O33631;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE M-LIKE PROTEIN.  
OS Streptococcus equisimilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=119602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.equisimilis; STRAIN=25287;  
RA Geyer A., Schmidt K.H.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.dysgalactiae subsp. equisimilis; STRAIN=25287;  
RX MEDLINE=20123454; PubMed=10660058;  
RA Geyer A., Schmidt K.H.;  
RT "Genetic organisation of the M protein region in human isolates of  
group C and G streptococci: two types of multigene regulator-like  
(mgrc) regions.";  
RL Mol. Gen. Genet. 262:965-974(2000).  
DR EMBL; X93464; CAA63750.1; -.  
DR EMBL; Y18363; CAB70608.1; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003345; M\_repeat.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02370; M; 10.  
DR PRINTS; PS00015; GPOSANCHOR.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 475 AA; 52399 MW; DC077E86DA51EC8E CRC64;

Query Match 62.9%; Score 39; DB 2; Length 475;

Best Local Similarity 75.0%; Pred. No. 3.3e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAK 12  
DB 383 KLAKQAEELAKL 394  
: | | | | | | | |

RESULT 14  
Q55098  
ID Q55098 PRELIMINARY; PRT; 532 AA.  
AC Q55098;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE M PROTEIN.  
GN EMM.  
OS Streptococcus sp.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GROUP C, C74A (ATCC 1238);  
RA Podbielski A., Melzer B.;  
RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X60097; CAA42693.1; -.  
DR HSP; P03069; IGCL.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003345; M\_repeat.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR Pfam: PF02370; M; 8.  
DR PRINTS: PR00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 532 AA; 60362 MW; 4A1A78B9D1122787 CRC64;

Query Match 62.9%; Score 39; DB 2; Length 532;  
Best Local Similarity 75.0%; Pred. No. 3.7e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKL 12  
| | | | | : | | | |  
Db 440 KLAKQAEELAKL 451

RESULT 15  
Q54840  
ID Q54840 PRELIMINARY; PRT: 550 AA.  
AC Q54840;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE M PROTEIN.  
GN EMM55.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID-1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SEROTYPE M55;  
RX MEDLINE-95371690; PubMed-7643859;  
RA Boyle M.D.P., Weber-Heynemann J., Raeder R., Podbielski A.;  
RT "Characterization of a gene coding for a type Ito bacterial IgG-binding protein.";  
RL Mol. Immunol. 32:669-678(1995).  
DR EMBL; X72090; CAA50980.1; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF0746; Gram\_pos\_anchor; 1.  
DR PRINTS: PR00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 550 AA; 61736 MW; 32894BFF9805181B CRC64;

Query Match 62.9%; Score 39; DB 2; Length 550;  
Best Local Similarity 75.0%; Pred. No. 3.8e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKL 12  
| | | | | : | | | |  
Db 459 KLAKQAEELAKL 470

Search completed: May 8, 2002, 07:20:14  
Job time: 142 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:17:52 ; Search time 12.86 seconds  
(without alignments)  
39.915 Million cell updates/sec

Title: US-09-765-086-200  
Perfect score: 62  
Sequence: 1 KLAKLAKLAKLAK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	43	69.4	1722	1	RB82_HUMAN		P29375 homo sapien
2	39	62.9	254	1	YCAL_ECOLI		P43674 escherichia
3	39	62.9	539	1	M24_STRPY		P12379 streptococc
4	38	61.3	142	1	ATPE_HAEIN		P43718 haemophilus
5	38	61.3	217	1	H1G_STRPY		P07796 strongyloce
6	38	61.3	279	1	YJ86_YEAST		P47153 saccharomyc
7	38	61.3	439	1	Y811_METJA		Q58221 methanococc
8	38	61.3	577	1	ORC2_HUMAN		Q13416 homo sapien
9	38	61.3	1391	1	MST2_DROHY		Q08696 drosophila
10	37	59.7	112	1	HMGD_DROME		Q05783 drosophila
11	37	59.7	340	1	Y269_MYCGE		Q49407 mycoplasma
12	37	59.7	424	1	GSA_CAMEJ		Q99870 campylobact
13	37	59.7	601	1	CORO_SCHPO		O13923 schizosacch
14	37	59.7	810	1	PLSB_HAEIN		P44857 haemophilus
15	37	59.7	836	1	NOY3_YEAST		P06102 saccharomyc
16	37	59.7	951	1	SVF_ECOLI		P07118 escherichia
17	36	58.1	92	1	RPC1_BPP22		P03041 bacterioph
18	36	58.1	171	1	FAB3_CAEEL		Q20222 caenorhabdi
19	36	58.1	219	1	H1B_XENLA		P06893 xenopus lae
20	36	58.1	259	1	RPOD_PYRHO		O59303 pyrococcus
21	36	58.1	452	1	TRPC_HELPJ		Q92308 helicobacte
22	36	58.1	485	1	SAHL_MESCR		P93253 mesembryant
23	36	58.1	485	1	SAHL_PETCR		Q01781 petroselinu
24	36	58.1	485	1	SAHL_TOBAC		P50248 nicotiana t
25	36	58.1	713	1	NU82_YEAST		P40368 saccharomyc
26	36	58.1	720	1	SPOT_MYCGE		P47520 mycoplasma
27	36	58.1	800	1	RR3_CHLEU		Q93324 chlamydomon
28	36	58.1	900	1	MANE_CAEEL		Q93324 caenorhabdi
29	36	58.1	1738	1	YCF1_EPIVI		Q00383 epifagus vi
30	36	58.1	2479	1	POLN_RRVN		P13887 ross river
31	35	56.5	189	1	KTHI_SULSO		Q9UXG7 sulfolobus
32	35	56.5	344	1	Y240_MYCPN		P75442 mycoplasma
33	35	56.5	376	1	SPH_STRPY		P50470 streptococc

34	35	56.5	386	1	ARP4_STRPY	P13050 streptococc
35	35	56.5	388	1	MRP4_STRPY	P30141 streptococc
36	35	56.5	388	1	PAM_STRPY	P49054 streptococc
37	35	56.5	389	1	M49_STRPY	P16947 streptococc
38	35	56.5	407	1	M21_STRPY	P50468 streptococc
39	35	56.5	439	1	Y412_ARATH	O04658 arabidopsis
40	35	56.5	480	1	PGKH_WHEAT	P12782 triticum ae
41	35	56.5	482	1	T2EA_YEAST	P36100 saccharomyc
42	35	56.5	483	1	M6_STRPY	P08089 streptococc
43	35	56.5	485	1	SAHL_CATRO	P35007 catharanthu
44	35	56.5	489	1	YABN_BACSU	P37556 bacillus su
45	35	56.5	492	1	M5_STRPY	P02977 streptococc

ALIGNMENTS

RESULT 1

RB82_HUMAN	RB82_HUMAN	STANDARD;	PRT;	1722 AA.
AC	P29375;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	RETINOBLASTOMA BINDING PROTEIN 2 (RBBP-2).			
GN	RBBP2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94020841; PubMed=8414517;			
RA	Fattaeey A.R., Helin K., Dembski M.S., Dyson N., Harlow E.,			
RA	Vuocolo G.A., Hanobik M.G., Haskell K.M., Oliff A., Defeo-Jones D.,			
RA	Jones R.E.;			
RT	"Characterization of the retinoblastoma binding proteins RBP1 and			
RT	RBP2.";			
RL	Oncogene 8:3149-3156(1993).			
RN	[2]			
RP	SEQUENCE OF 1102-1569 FROM N.A.			
RX	MEDLINE=91312450; PubMed=1857421;			
RA	Defeo-Jones D., Huang P.S., Jones R.E., Haskell K.M., Vuocolo G.A.,			
RA	Hanobik M.G., Huber H.E., Oliff A.;			
RT	"Cloning of cDNAs for cellular proteins that bind to the			
RT	retinoblastoma gene product.";			
RL	Nature 352:251-254(1991).			
CC	-1- FUNCTION: INTERACTS WITH THE VIRAL PROTEIN-BINDING DOMAIN OF THE			
CC	RETINOBLASTOMA PROTEIN.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).			
CC	-1- SIMILARITY: TO THE C-TERMINAL DOMAIN OF THE XE169 PROTEIN AND			
CC	THE JUMONJI PROTEIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; S66431; AAB28544.1; -			
DR	PIR; S16954; S16954.			
DR	MIM; 180202; -			
DR	InterPro; IPR001606; ARID.			
DR	InterPro; IPR003347; JmJC.			
DR	InterPro; IPR003349; JmJN.			
DR	InterPro; IPR001965; PHD.			
DR	Pfam; PF01388; ARID; 1.			
DR	Pfam; PF02373; JmJC; 1.			
DR	Pfam; PF02375; JmJN; 1.			
DR	Pfam; PF00628; PHD; 3.			
DR	SMART; SM00501; BRIGHT; 1.			

```

DR SMART; SM00545; JmJN; 1.
DR SMART; SM00249; PHD; 3.
KW Trans-acting factor; Nuclear protein.
FT CONFLICT 1563 1566 MISSING (IN REF. 2).
SQ SEQUENCE 1722 AA; 195815 MW; 8CFF8A88AE69A652 CRC64;

Query Match 69.4%; Score 43; DB 1; Length 1722;
Best Local Similarity 90.9%; Pred. No. 40;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAK 11
Db 1552 KLKLAKLAKLAK 1562

RESULT 2
YCAL_ECOLI
ID YCAL_ECOLI STANDARD; PRT; 254 AA.
AC P43674; P75840;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE METALLOPROTEASE YCAL (EC 3.4.24.-).
GN YCAL OR B0909.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 1-183 FROM N.A.
RC STRAIN-K12 / MG1655;
RA Palma C.A., Allen E., Araujo R., Aparicio A.M., Botstein D.,
RA Cherry M., Chung E., Dietrich F., Duncan M., Federspiel N.,
RA Kalman S., Kim K., Komp C., Lashkari D., Lew H., Lin D.,
RA Namath A., Oefner P., Davis R.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 150-254 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=95138007; PubMed=7836281;
RA Fricke J., Neuhaed J., Kelln R.A., Pedersen S.;
RT "The cna gene encoding cytidine monophosphate kinase is located in
RT the rpsA operon and is required for normal replication rate in
RT Escherichia coli.";
RL J. Bacteriol. 177:517-523(1995).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48 (ZINC
CC METALLOPROTEASE). SPONG, TO E.COLI YGGG.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC EMBL; M19031; AAA26874.1; -.
DR PIR; A28549; A28549.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02370; M; 11.
DR PRINTS; PR0015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Signal; Virulence; Cell wall; Transmembrane; Phagocytosis;
KW Duplication; Repeat; Antigen; Coiled coil.
FT SIGNAL 1 42 POTENTIAL.
FT CHAIN 43 539 M PROTEIN, SEROTYPE 24.
FT DOMAIN 43 514 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 515 534 MEMBRANE ANCHOR.
FT DOMAIN 535 539 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 118 301 5.3 X 35 AA TANDEM REPEATS, A-TYPE.
FT REPEAT 118 152 A-1.
FT REPEAT 153 187 A-2 (1 ALTERATION).
FT REPEAT 188 222 A-3 (2 ALTERATIONS).
FT REPEAT 223 257 A-4 (1 ALTERATION).
FT REPEAT 258 292 A-5 (2 ALTERATIONS).
FT REPEAT 293 301 A-6 (INCOMPLETE).
FT DOMAIN 311 405 2.7 X 35 AA TANDEM REPEATS, B-TYPE.
FT REPEAT 311 355 B-1.
FT REPEAT 356 380 B-2.
FT REPEAT 381 405 B-3 (INCOMPLETE).
FT DOMAIN 468 504 GLY/PRO-RICH (CELL WALL-SPANNING). SURFACE
FT DOMAIN 505 510 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 539 AA; 58804 MW; B03EDF3AC1B5E9C7 CRC64;

```

```

Query Match 62.9%; Score 39; DB 1; Length 539;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 KLAKLAKKLAKL 12
    |||||:|||||
DB 447 KLAKAEELAKL 458

```

```

RESULT 4
ATPE_HAEIN
ID ATPE_HAEIN STANDARD; PRT; 142 AA.
AC P43718;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).
GN ATPC OR H10478.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kertavagae A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.

```

```

CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32730; AAC22136.1; -.
CC HSSP; P00832; 1BSH.
CC TIGR; H10478; -.
CC InterPro; IPR001469; ATP-synt_DE.
CC Pfam; PF00401; ATP-synt_DE; 1.
CC ProDom; PD000944; ATP-synt_DE; 1.
KW Hydrolase; ATP synthesis; CF(1); Hydrogen ion transport;
KW Complete proteome.
SQ SEQUENCE 142 AA; 15591 MW; E5BDE1DA34A9039 CRC64;

Query Match 61.3%; Score 38; DB 1; Length 142;
Best Local Similarity 72.7%; Pred. No. 25;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LAKLAKKLAKL 12
    |||||:|||||
DB 118 VAKLSKELAKL 128

RESULT 5
HIG_STRPU
ID HIG_STRPU STANDARD; PRT; 217 AA.
AC P07796;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H1-GAMMA, LATE.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OC NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87172742; PubMed=3031476;
RA Knowles J.A., Lai Z.-C., Childs G.J.;
RT "Isolation, characterization, and expression of the gene encoding the
RT late histone subtype H1-gamma of the sea urchin Strongylocentrotus
RT purpuratus."
RL Mol. Cell. Biol. 7:478-485(1987).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16033; AAA30059.1; -.
CC PIR; A26721; A26721.
CC HSSP; P02259; IHST.
CC InterPro; IPR001386; Linker_histone.
CC InterPro; IPR003216; Linkerhist_N.

```

```
DR Pfam: PF00538; linker_histone; 1.
DR ProDom: PD000373; Linkerhist_N; 1.
DR SMART: SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 217 AA; 22658 MW; C725IED3413B185 CRC64;

Query Match 61.3%; Score 38; DB 1; Length 217;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KLAKLAKKLAKLAK 14
   | | | | | | | |
Db 196 KAAKPAKKAAPAK 209

RESULT 6
YJ86_YEAST
ID YJ86_YEAST STANDARD; PRT; 279 AA.
AC P47153;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 32.0 KDA PROTEIN IN NNFI-SFE24 INTERGENIC REGION.
GN YJR116W OR J2031.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RC Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49616; CRA89046.1; -.
DR SGD: S0003877; YJR116W.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
SQ SEQUENCE 279 AA; 32011 MW; 66102933B38B7F2B5.CRC64;

Query Match 61.3%; Score 38; DB 1; Length 279;
Best Local Similarity 61.5%; Pred. No. 45;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KLAKLAKKLAKLA 13
   | : : | | | | |
Db 261 KMIRIAKLAKPA 273

RESULT 7
YB11_METJA
ID YB11_METJA STANDARD; PRT; 439 AA.
AC Q58221;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0811.
```

```
GN MJ0811.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO M.JANNASCHII MJ0077 AND SOME, TO E.COLI YIEM.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67525; AAB98810.1; -.
DR HSSP: P25230; ILYP.
DR TIGR: MJ0811.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWFA; 1.
DR PROSITE: PS0234; VWFA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 273 439 VWFA.
SQ SEQUENCE 439 AA; 51119 MW; 2F85F2B942C1CDFD CRC64;

Query Match 61.3%; Score 38; DB 1; Length 439;
Best Local Similarity 69.2%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KLAKLAKKLAKLA 13
   | | | | | | | |
Db 111 KLAKLKKELKKFA 123

RESULT 8
ORC2_HUMAN
ID ORC2_HUMAN STANDARD; PRT; 577 AA.
AC Q13416; Q13204;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ORIGIN RECOGNITION COMPLEX SUBUNIT 2.
GN ORC2L OR ORC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96099401; PubMed=7502077;
RA Gavin K.A., Hidaka M., Stillman B.D.;
RT "Conserved initiator proteins in eukaryotes."
RL Science 270:1667-1671(1995).
RN [2]
RP REVISIONS TO 41-42.
RA Hidaka M., Stillman B.;
```



Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RL [3]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=96404011; PubMed=8808289;  
RX Takahara K., Bong M., Breward R., Eddy R.L., Haley L.L., Sait S.J.,  
RA Shows T.B., Hoffman G.G., Greenspan D.S.;  
RT "Mouse and human homologues of the yeast origin of replication  
RT recognition complex subunit ORC2 and chromosomal localization of the  
RT cognate human gene ORC2L";  
RL Genomics 31:119-122(1996).

CC -!- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT  
CC BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL  
CC REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO  
CC THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN  
CC ATP-DEPENDENT MANNER.  
CC -!- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: BELONGS TO THE ORC2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U40268; AAC50326.2; -;  
DR EMBL: U27459; AAB33970.1; -;  
DR MIM: 601182; -;  
KW DNA replication; Nuclear protein.  
FT CONFLICT 131 131 I -> V (IN REF. 3).  
FT CONFLICT 236 236 T -> L (IN REF. 3).  
FT CONFLICT 392 392 I -> M (IN REF. 3).  
SQ SEQUENCE 577 AA; 65971 MW; DF3F9C2CF147DA5F CRC64;

Query Match 61.3%; Score 38; DB 1; Length 577;  
Best Local Similarity 72.7%; Pred. No. 85;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 KLAKLAKKLAK 11  
|:||||:||||  
Db 105 KMAKLASELAK 115

RESULT 9  
MST2\_DROHY ID MST2\_DROHY STANDARD; PRT; 1391 AA.  
AC Q08696;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE AXONEME-ASSOCIATED PROTEIN MST101(2).  
GN MST101(2).  
OS Drosophila hydei (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7224;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=95045538; PubMed=7957199;  
RA Neesen J., Padmanabhan S., Buennenmann H.;  
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid  
RT motif representing the major component of the sperm-tail-specific  
RT axoneme-associated protein family Dhmst101 form extended  
RT alpha-helical rods within the extremely elongated spermatozoa of  
RT Drosophila hydei";  
RL Eur. J. Biochem. 225:1089-1095(1994).  
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY

CC SPERMATIDS.  
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.  
CC -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT  
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM  
CC REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X73481; CAA51876.1; -;  
DR PIR: S34154; S34154.  
DR HSSP: P01032; IC5A.  
DR FlyBase; FBgn0020733; Dhyd\mst101(2).  
KW Sperm; Repeat; Multigene family; Polymorphism.  
FT DOMAIN 332 1268 [KRI-K-X-C-X-X-A-K-X-X-X-X-E.  
FT SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

Query Match 61.3%; Score 38; DB 1; Length 1391;  
Best Local Similarity 60.0%; Pred. No. 1.8e+02;  
Matches 12; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

OY 1 KLAKLA-----KKLAKLAK 14  
|:||||:|  
Db 1319 KLAKKAKEAIEWKKCAKLAK 1338

RESULT 10  
HMGD\_DROME ID HMGD\_DROME STANDARD; PRT; 112 AA.  
AC Q05783; Q9W2D3;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HIGH MOBILITY GROUP PROTEIN D (HMG-D).  
GN HMGD OR CG17950.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 45-77.  
RC TISSUE=Embryo;  
RX MEDLINE=92236564; PubMed=1373803;  
RA Wagner C.R., Hamana K., Elgin S.C.R.;  
RT "A high-mobility-group protein and its cDNAs from Drosophila  
RT melanogaster";  
RL Mol. Cell. Biol. 12:1915-1923(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CANTON-S;  
RX MEDLINE=94021387; PubMed=8414994;  
RA Ner S.S., Churchill M.E.A., Searles M.A., Travers A.A.;  
RT "dHMG-Z, a second HMG-1-related protein in Drosophila melanogaster";  
RL Nucleic Acids Res. 21:4369-4371(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RN CHARACTERIZATION.  
RX MEDLINE-94222028; PubMed-8168480;  
RA Ner S.S., Travers A.A.;  
RT "HMG-D, the Drosophila melanogaster homologue of HMG 1 protein, is  
RT associated with early embryonic chromatin in the absence of histone  
RT H1.";  
RL EMBO J. 13:1817-1822(1994).  
RN [5]  
RN CHARACTERIZATION.  
RX MEDLINE-95237208; PubMed-7720717;  
RA Churchill M.E.A., Jones D.N.M., Glaser T., Hefner H., Searles M.A.,  
RA Travers A.A.;  
RT "HMG-D is an architecture-specific protein that preferentially binds  
RT to DNA containing the dinucleotide TG.";  
RL EMBO J. 14:1264-1275(1995).  
RN [6]  
RN STRUCTURE BY NMR OF 2-74.  
RX MEDLINE-95006330; PubMed-7922039;  
RA Jones D.N.M., Searles M.A., Shaw G.L., Churchill M.E.A., Ner S.S.,  
RA Keeler J., Travers A.A., Neuhaus D.;  
RT "The solution structure and dynamics of the DNA-binding domain of  
RT HMG-D from Drosophila melanogaster.";  
RL Structure 2:609-627(1994).  
CC -!- FUNCTION: BINDS PREFERENTIALLY SINGLE-STRANDED DNA AND UNWINDS  
CC DOUBLE STRANDED DNA. PREFERENCES SITES CONTAINING THE SEQUENCE 5'-ATG-  
CC 3'. FACILITATES DNA BENDING. ASSOCIATED WITH EARLY EMBRYONIC  
CC CHROMATIN IN THE ABSENCE OF HISTONE H1.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES OF DEVELOPMENT.  
CC -!- SIMILARITY: BELONGS TO THE HMG1/HMG2 PROTEIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X71138; CAA50468.1; -

DR EMBL; M77023; AAA28609.1; -  
DR EMBL; AE003455; AAF46759.1; -  
DR PIR; A44382; A44382.  
DR PIR; S32724; S32724.  
DR PDB; 1HMA; 31-JUL-94.  
DR FlyBase; FBgn0004362; HmgD.  
DR InterPro; IPR000910; HMG\_12\_box.  
DR Pfam; PF00505; HMG\_box; 1.  
DR SMART; SM00398; HMG; 1.  
KW Nuclear protein; Chromosomal protein; DNA-binding; 3D-structure.  
FT DNA\_BIND 5 71 HMG BOX.  
FT DOMAIN 101 112 ASP/GLU-RICH (ACIDIC).  
SQ SEQUENCE 112 AA; 12416 MW; 3F537CCFD62FEC9F CRC64;  
  
Query Match 59.7%; Score 37; DB 1; Length 112;  
Best Local Similarity 64.3%; Pred No. 29;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 KIATLAKLAKLAKLAK 14  
| | | | | : |  
DB 86 KRKPAKKVAKKSK 99  
  
RESULT 11  
Y269 MYCGE  
ID Y269 MYCGE STANDARD; PRT; 340 AA.  
AC Q9407; Q49328;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN MG269.  
GN MG269.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE-96026346; PubMed-7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
RN [2]  
RP SEQUENCE OF 1-79 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE-94075230; PubMed-8253680;  
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
RT sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U39706; AAC71491.1; -  
DR EMBL; U02215; AAD12510.1; ALT\_INIT.  
DR TIGR; MG269; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 340 AA; 39662 MW; 3921FELF81501CC4 CRC64;

Query Match 59.7%; Score 37; DB 1; Length 340;  
Best Local Similarity 58.3%; Pred. No. 76;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AKLAKKLAKLAK 14  
||| |||::: |  
Db 271 AKLTKIAETK 282

## RESULT 12

GSA\_CAMJE STANDARD; PRT; 424 AA.  
AC Q9PP70;  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GLUTAMATE-1-SEMIALDEHYDE 2,1-AMINOMUTASE (EC 5.4.3.8) (GSA)  
DE (GLUTAMATE-1-SEMIALDEHYDE AMINOTRANSFERASE) (GSA-AT).  
GN HEML OR CJ0853.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jørgels K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vleet A.H.M.,  
RA Whitehead S., Barrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
CC -1- CATALYTIC ACTIVITY: (S)-4-AMINO-5-OXOPENTANOATE -> 5-  
CC AMINOLEVULINATE.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY; SECOND STEP.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: AL139076; CAB73118.1; -;  
DR InterPro: IPR000954; AminoTran\_3.  
DR Pfam: PF00202; aminotran\_3; 1.  
DR PROSITE: PS00600; AA\_TRANSFER\_CLASS\_3; FALSE\_NEG.  
KW Porphyrin biosynthesis; isomerase; pyridoxal phosphate;  
KW Complete proteome.  
FT BINDING 263 263 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 424 AA; 46092 MW; 7811526A5AEDA57D CRC64;

Query Match 59.7%; Score 37; DB 1; Length 424;  
Best Local Similarity 72.7%; Pred. No. 92;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KLAKLAKLAK 11  
||| |||::: |  
Db 322 KLGLAKLKTQ 332

## RESULT 13

CORO\_SCHPO STANDARD; PRT; 601 AA.  
AC O13923;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE CORONIN-LIKE PROTEIN CRN1.  
DE CORONIN-LIKE PROTEIN CRN1.  
GN CRN1 OR SPAC23C4.02.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RL Harris D., Squares R., Barrell B.G., Rajandream M.A., Wood V.;  
Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: BINDS TO F-ACTIN (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
CC -1- SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: Z99753; CAB16873.1; -;  
DR InterPro: IPR001680; WD40.  
DR Pfam: PF00400; WD40; 3.  
DR PRINTS: PR00320; GPROTEINBRPT.  
DR SMART: SM00320; WD40; 3.  
DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
DR PROSITE: PS00082; WD\_REPEATS\_2; 3.  
DR PROSITE: PS0294; WD\_REPEATS\_REGION; 1.  
KW Actin-binding; Repeat; WD repeat; Coiled coil.  
FT REPEAT 79 119 WD 1.  
FT REPEAT 132 172 WD 2.  
FT REPEAT 174 213 WD 3.  
FT REPEAT 220 260 WD 4.  
FT REPEAT 266 306 WD 5.  
FT DOMAIN 556 600 COILED COIL (POTENTIAL).  
SQ SEQUENCE 601 AA; 67015 MW; 231096AE76CADE4D CRC64;

Query Match 59.7%; Score 37; DB 1; Length 601;  
Best Local Similarity 57.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAK 14  
|||::| ||||| |  
Db 581 KIAQEDKLAKITE 594

## RESULT 14

PLSB\_HAEIN STANDARD; PRT; 810 AA.  
AC P44857;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (EC 2.3.1.15) (GPAT).  
DE PLSB OR HI0748.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;

```

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervilave A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ACYL-COA + GLYCEROL 3-PHOSPHATE = COA +
CC 1-ACYL-GLYCEROL 3-PHOSPHATE.
CC -1- PATHWAY: DE NOVO PHOSPHOLIPID BIOSYNTHESIS: FIRST STEP. MAY ALSO
CC FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32758; AAC22406.1; -
DR TIGR: H10748; -
DR InterPro: IPR002123; Acyltransferase.
DR Pfam: PF01553; Acyltransferase; 1.
DR K W Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
KW Complete proteome.
SQ SEQUENCE 810 AA; 92774 MW; DB8564BC3E2C370D CRC64;
-----
Query Match 59.7%; Score 37; DB 1; Length 810;
Best Local Similarity 63.6%; Pred. No. 1.6e+02;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 4 KLAKLAKLAK 14
Db 185 KLAQLARAK 195
|:|:|:|:|:|

RESULT 15
NOT3_YEAST
ID NOT3_YEAST STANDARD; PRT; 836 AA.
AC P06102;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 3.
GN NOT3 OR YIL038C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86312926; PubMed=3018676;
RA Ferguson J., Ho J.-Y., Peterson T.A., Reed S.I.;
RT "Nucleotide sequence of the yeast cell division cycle start genes
RT CDC28, CDC36, CDC37, and CDC39, and a structural analysis of the
RT predicted products.";
RL Nucleic Acids Res. 14:6681-6697(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,

```

```

RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=95011559; PubMed=7926748;
RA Collart M.A., Struhl K.;
RT "NOT1(CDC39), NOT2(CDC36), NOT3, and NOT4 encode a global-negative
RT regulator of transcription that differentially affects TATA-element
RT utilization.";
RL Genes Dev. 8:525-537(1994).
CC -1- FUNCTION: NEGATIVELY REGULATES THE BASAL AND ACTIVATED
CC TRANSCRIPTION OF MANY GENES. PREFERENTIALLY AFFECTS TC-TYPE TATA
CC ELEMENT-DEPENDENT TRANSCRIPTION. COULD DIRECTLY OR INDIRECTLY
CC INHIBITS COMPONENT(S) OF THE GENERAL TRANSCRIPTION MACHINERY.
CC -1- SUBUNIT: FORMS A COMPLEX THAT COMPRISE NOT1, NOT2, NOT3, NOT4 AND
CC NOT5.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TO YEAST NOT5.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE CDC29 (WHICH IS IN
CC FACT NOT1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z46861; CAA86913.1; -
DR EMBL: X04289; CAA27837.1; -
DR PIR: C26372; C26372.
DR SGD: S0001300; NOT3.
KW Nuclear protein; Transcription regulation; Repressor; Coiled coil.
FT DOMAIN 36 68 COILED COIL (POTENTIAL).
FT DOMAIN 119 195 COILED COIL (POTENTIAL).
FT DOMAIN 255 292 COILED COIL (POTENTIAL).
FT DOMAIN 803 831 COILED COIL (POTENTIAL).
FT CONFLICT 106 106 I -> N (IN REF. 1).
FT CONFLICT 725 725 Y -> C (IN REF. 1).
FT CONFLICT 827 836 KOLKOGKISV -> ETIETGKN (IN REF. 1).
SQ SEQUENCE 836 AA; 94402 MW; 69B76694FCC6846F CRC64;
-----
Query Match 59.7%; Score 37; DB 1; Length 836;
Best Local Similarity 69.2%; Pred. No. 1.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKIA 13
Db 276 KLREAKKAKIA 288
|:|:|:|:|:|

Search completed: May 8, 2002, 07:20:33
Job time: 161 sec

```

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:17:52 ; Search time 22.68 seconds  
(without alignments)  
47.021 Million cell updates/sec

Title: US-09-765-086-200  
Perfect score: 62  
Sequence: 1 KLAKLAKKLAK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	69.4	1722	1 I78879	retinoblastoma bin
2	40.5	65.3	476	2 B96667	hypothetical prote
3	40	64.5	222	2 D81122	baseplate assembly
4	40	64.5	426	2 D72252	conserved hypotet
5	39.5	63.7	648	2 T23864	hypothetical prote
6	39	62.9	262	2 F85616	probable heat choc
7	39	62.9	262	2 D64830	cmk protein precu
8	39	62.9	436	2 S30284	M protein precu
9	39	62.9	454	2 S43556	plasmaingogen-bindin
10	39	62.9	455	2 D81284	probable fumarate
11	39	62.9	472	2 S43554	plasmaingogen-bindin
12	39	62.9	532	2 S54871	M protein - Strept
13	39	62.9	539	2 A28549	M24 protein precu
14	39	62.9	587	2 JC1419	Fc gamma (IgG) rec
15	38	61.3	142	2 C64071	H+-transporting At
16	38	61.3	217	2 A26721	histone H1-gamma,
17	38	61.3	279	2 S57139	probable membrane
18	38	61.3	439	2 C64401	hypothetical prote
19	38	61.3	461	2 S54543	probable membrane
20	38	61.3	462	2 T15543	hypothetical prote
21	38	61.3	1184	2 T41515	coiled coil protei
22	38	61.3	1390	2 S51364	sperm tail-specifi
23	37	59.7	112	2 A44382	high mobility grou
24	37	59.7	127	2 B72593	probable gastrin p
25	37	59.7	246	2 T21215	hypothetical prote
26	37	59.7	282	2 C84114	UDP-glucose 4-epim
27	37	59.7	340	2 G64229	surface protein an
28	37	59.7	347	2 A81794	probable rotamase
29	37	59.7	393	2 C64466	GTP-binding protei

30 37 59.7 412 2 T16480  
31 37 59.7 424 2 E81358  
32 37 59.7 537 2 T39037  
33 37 59.7 601 2 T38258  
34 37 59.7 634 2 T27465  
35 37 59.7 758 2 H75013  
36 37 59.7 810 2 D64090  
37 37 59.7 836 2 S49940  
38 37 59.7 951 1 SYECVT  
39 37 59.7 951 2 E86124  
40 36 58.1 92 1 Z1BPC2  
41 36 58.1 218 2 H70371  
42 36 58.1 219 1 HSX11B  
43 36 58.1 220 2 I51447  
44 36 58.1 227 2 I54251  
45 36 58.1 227 2 T15035

hypothetical prote  
glutamate-1-semial  
hypothetical gtp-b  
coronin-like prote  
hypothetical prote  
hypothetical prote  
glycerol-3-phospha  
cell division cont  
valine--trNA ligas  
valine tRNA synthe  
regulatory protein  
trans-regulatory p  
histone H1B - Afri  
histone H1B - Afri  
translation elonga  
adenosylhomocyste

## ALIGNMENTS

## RESULT 1

I78879

retinoblastoma binding protein 2 - human

C:Species: Homo sapiens (man)

C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 17-Mar-2000

C:Accession: I78879; S16954

R:Fattae, A.R.; Helin, K.; Dembski, M.S.; Dyson, N.; Harlow, E.; Vuocolo, G.A.; Hano

Oncogene 8, 3149-3156, 1993

A:Title: Characterization of the retinoblastoma binding proteins RBP1 and RBP2.

A:Reference number: I58383; MUID:94020841

A:Accession: I78879

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1722 &lt;FAT&gt;

R:Cross-references: GB:S66431; NID:g435777; PIDN:AAB28544.1; PID:g435778

R:Defeo-Jones, D.; Huang, P.S.; Jones, R.E.; Haskell, K.M.; Vuocolo, G.A.; Hanobik, M.

Nature 352, 251-254, 1991

A:Title: Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene

A:Reference number: S16953; MUID:91312450

A:Accession: S16954

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1102-1562, 'KKK' &lt;DEF&gt;

C:Genetics:

A:Gene: GDB:RBP2

A:Cross-references: GDB:119548; OMIM:180280

A:Map position: 3q21-3qter

C:Superfamily: human retinoblastoma binding protein 2

Query Match 69.4%; Score 43; DB 1; Length 1722;  
Best Local Similarity 90.9%; Pred. No. 82;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAK 11

|| |||||

Db 1552 KLNLAKKLAK 1562

## RESULT 2

B96667

hypothetical protein F15H21.20 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: B96667

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719

A:Accession: B96667

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <STO>

A:Cross-references: GB:AE005173; NID:g10645403; PIDN:AAG21521.1; GSPDB:GN00141

C:Genetics:

A:Gene: F15H21.20

A:Map position: 1

Query Match 65.3%; Score 40.5; DB 2; Length 476;

Best Local Similarity 57.1%; Pred. No. 62;

Matches 12; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

QY 1 KLAKLAKKLA-----KLAK 14

DB 393 KLSKLEKLAEBGTEKCLKAK 413

RESULT 3

D81122  
baseplate assembly protein V, probable NMB1111 [imported] - *Neisseria meningitidis* (strain D81122)

C:Species: *Neisseria meningitidis*

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Feb-2001

C:Accession: D81122; A81901

R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

Rick, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter

A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: D81122

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-222 <TEP>

A:Cross-references: GB:AE002460; GB:AE002098; NID:g7226335; PIDN:AAFA1502.1; PID:g722634

A:Experimental source: serogroup B, strain MC58

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

et al.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.

A:Reference number: A81775; MUID:20225556

A:Accession: A81901

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-222 <PAR>

A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84572.1; PID:g737999

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMB1111; NMA1321

Query Match

Best Local Similarity 64.5%; Score 40; DB 2; Length 222;

Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LAKLAKKLA-----KLAK 14

DB 3 LSKLAKKTAQTAK 15

RESULT 4

D72252

conserved hypothetical protein - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: D72252

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: D72252

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-426 <ARN>

A:Cross-references: GB:AE001797; GB:AE000512; NID:g4982017; PIDN:AAD36535.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1467

Query Match 64.5%; Score 40; DB 2; Length 426;

Best Local Similarity 72.7%; Pred. No. 66;

Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKLAK 11

DB 260 KVAELAKKIAK 270

RESULT 5

T23864

hypothetical protein F39D8.4 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T23864

R:Hembry, C.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19809

A:Accession: T23864

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-648 <WIL>

A:Cross-references: EMBL:Z69793; PIDN:CAA93676.1; GSPDB:GN000028; CESP:F39D8.4

A:Experimental source: clone R03A10

C:Genetics:

A:Gene: CESP:F39D8.4

A:Map position: X

A:Introns: 22/2; 112/3; 160/1; 250/1; 334/1; 424/3; 463/1; 485/3; 528/2; 593/2

Query Match 63.7%; Score 39.5; DB 2; Length 648;

Best Local Similarity 73.3%; Pred. No. 1.1e+02;

Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KLAKLAKLAK-LAK 14

DB 376 KLAKLQKLGKTLAK 390

RESULT 6

F85616

probable heat shock protein ycaL [imported] - *Escherichia coli* (strain O157:H7)

C:Species: *Escherichia coli*

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: F85616

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

killer, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85616

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-262 <STO>

A:Cross-references: GB:AE005174; NID:g12514075; PIDN:AAG55394.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ycaL



Query Match 62.9%; Score 39; DB 2; Length 262;  
Best Local Similarity 57.1%; Pred. No. 61;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
I : I I : I I I : I I  
Db 77 KSKYKRLAKIAK 90

## RESULT 7

D64830  
cmk protein precursor - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C:Accession: D64830; I41019; S49998  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: D64830  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <BLAT>  
A:Cross-references: GB:AE000193; GB:U00096; NID:gl787134; PIDN:AACT3995.1; PID:gl787138;  
A:Experimental source: strain K-12, substrain MG1655  
R:Fricke, J.; Neuhaed, J.; Kelln, R.A.; Pedersen, S.  
J. Bacteriol. 177, 517-523, 1995  
A:Title: The cmk gene encoding cytidine monophosphate kinase is located in the rpsA operon  
A:Reference number: I41019; MUID:95138007  
A:Accession: I41019  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 158-248, 'DVRNFTSVIVSPVLSKLSLSLRWS', 251, 'SPL' <RES>  
A:Cross-references: EMBL:X82933; NID:g599733; PIDN:CAA58106.1; PID:g599734  
A:Experimental source: strain K-12; plasmid pSP547  
C:Genetics:  
A:Gene: cmk  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-262/Product: cmk protein #status predicted <MAT>

Query Match 62.9%; Score 39; DB 2; Length 262;  
Best Local Similarity 57.1%; Pred. No. 61;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
I : I I : I I I : I I  
Db 77 KSKYKRLAKIAK 90

## RESULT 8

S30284  
M protein precursor - Streptococcus pyogenes (serotype M52)  
C:Species: Streptococcus pyogenes  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 26-Aug-1999  
C:Accession: S30284; S29681  
R:Podbielski, A.  
Mol. Gen. Genet. 237, 287-300, 1993  
A:Title: Three different types of organization of the vir regulon in group A streptococci  
A:Reference number: S30283; MUID:93204905  
A:Accession: S30284  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-436 <PODI>  
A:Cross-references: EMBL:X58179  
R:Podbielski, A.; Welzer, B.  
Submitted to the EMBL Data Library, February 1991  
A:Description: Pcr mediated cloning and sequencing of group A streptococcal emm41/52 (ev  
A:Reference number: S29680  
A:Accession: S29681

A:Molecule type: DNA  
A:Residues: 1-216, 'N', 218-436 <POD2>  
A:Cross-references: EMBL:X58179; NID:g47364; PIDN:CAA41168.1; PID:g47365  
C:Genetics:  
A:Gene: emm  
C:Superfamily: M5 protein  
C:Keywords: transmembrane protein  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-436/Product: M protein #status predicted <MAT>  
F:411-429/Domain: transmembrane #status predicted <TMN>

Query Match 62.9%; Score 39; DB 2; Length 436;  
Best Local Similarity 75.0%; Pred. No. 96;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKL 12  
I I I I I : I I I I I  
Db 344 KLAKQAEELAKL 355

## RESULT 9

S43556  
plasmidogen-binding protein MLC36 - Streptococcus sp. (fragment)  
C:Species: Streptococcus sp.  
C:Date: 14-Sep-1994 #sequence\_revision 03-Nov-1995 #text\_change 26-Aug-1999  
C:Accession: S43558; S43556  
R:Ben Nasr, A.; Wistedt, A.; Ringdahl, U.; Sjoeborg, U.  
Eur. J. Biochem. 222, 267-276, 1994  
A:Title: Streptokinase activates plasminogen bound to human group C and G streptococci  
A:Reference number: S43558; MUID:94291620  
A:Accession: S43558  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <BE2>  
A:Cross-references: EMBL:Z32677; NID:g474767; PIDN:CAA83588.1; PID:g474768  
C:Superfamily: M5 protein

Query Match 62.9%; Score 39; DB 2; Length 454;  
Best Local Similarity 75.0%; Pred. No. 99;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKL 12  
I I I I I : I I I I I  
Db 389 KLAKQAEELAKL 400

## RESULT 10

D81284  
probable fumarate lyase Cj1394 [imported] - Campylobacter jejuni (strain NCTC 11168)  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: D81284  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
A:Reference number: AB1250; MUID:20150912  
A:Accession: D81284  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <PAR>  
A:Cross-references: GB:AL139078; GB:AL111168; NID:g6968723; PIDN:CAB73819.1; PID:g696  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: Cj1394  
C:Superfamily: fumarate hydratase

Query Match 62.9%; Score 39; DB 2; Length 455;  
Best Local Similarity 64.3%; Pred. No. 99;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAKL 14  
:| :|| :|||||  
Db 131 ELKATAKALAKLAK 144

## RESULT 11

S4354  
plasminogen-binding protein MLG72 - Streptococcus sp. (fragment)  
C:Species: Streptococcus sp.  
A:Variety: group G  
C:Date: 07-Sep-1994 #sequence\_revision 03-Nov-1995 #text\_change 20-Jun-2000  
C:Accession: S45599; S43554  
R:Ben Nasr, A.; Wistedt, A.; Ringdahl, U.; Sjoebbring, U.  
Eur. J. Biochem. 222, 267-276, 1994  
A:Title: Streptokinase activates plasminogen bound to human group C and G streptococci  
A:Reference number: S45598; MUID:94291620  
A:Accession: S45599  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-472 <BE2>  
A:Cross-references: EMBL:232678; NID:g474769; PIDN:CAA83589.1; PID:g1333838  
C:Superfamily: M5 protein

Query Match 62.9%; Score 39; DB 2; Length 472;

Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAKL 12  
||||| :|||  
Db 407 KLAQAELAKL 418

## RESULT 12

S54871  
M protein - Streptococcus sp.  
C:Species: Streptococcus sp.  
C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 26-Aug-1999  
C:Accession: S54871  
R:Podbielski, A.; Melzer, B.  
submitted to the ENBL Data Library, June 1991  
A:Reference number: S54871  
A:Accession: S54871  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-532 <POD>  
A:Cross-references: EMBL:X60097; NID:g840905; PIDN:CAA42693.1; PID:g840906  
C:Superfamily: M5 protein

Query Match 62.9%; Score 39; DB 2; Length 532;

Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAKL 12  
||||| :|||  
Db 440 KLAQAELAKL 451

## RESULT 13

A28549  
M24 protein precursor - Streptococcus pyogenes  
C:Species: Streptococcus pyogenes  
A:Variety: serotype M24  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 26-Aug-1999  
C:Accession: A28549; S60802  
R:Mouw, A.R.; Beachey, E.H.; Burdett, V.  
J. Bacteriol. 170, 676-684, 1988  
A:Title: Molecular evolution of streptococcal M protein: cloning and nucleotide sequence  
A:Reference number: A28549; MUID:88115166  
A:Accession: A28549  
A:Molecule type: DNA

A:Residues: 1-539 <MOU>  
A:Cross-references: GB:M19031; NID:g153616; PIDN:AAA26874.1; PID:g153617  
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.  
Mol. Microbiol. 14, 619-631, 1994  
A:Title: Non-congruent relationships between variation in emm gene sequences and the  
A:Reference number: S60784; MUID:95198337  
A:Accession: S60802  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual  
A:Molecule type: DNA  
A:Residues: 30-89 <WHA>  
C:Superfamily: M5 protein  
C:Keywords: coiled coil; transmembrane protein

Query Match 62.9%; Score 39; DB 2; Length 539;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAKL 12  
||||| :|||  
Db 447 KLAQAELAKL 458

## RESULT 14

JC1419  
Fc gamma (IgG) receptor II precursor - Streptococcus sp.  
N:Alternate names: fcrV protein  
C:Species: Streptococcus sp.  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 26-Aug-1999  
C:Accession: JC1419; SI7354  
R:Smirnov, O.Y.; Denesyuk, A.I.; Zakharov, M.V.; Abramov, V.M.; Zav'yalov, V.P.  
Gene 120, 27-32, 1992  
A:Title: Protein V, a novel type-II IgG receptor from Streptococcus sp.: Sequence, ho  
A:Reference number: JC1419; MUID:93013016  
A:Accession: JC1419  
A:Molecule type: DNA  
A:Residues: 1-587 <SMI>  
A:Cross-references: EMBL:X62467; NID:g47562; PIDN:CAA44324.1; PID:g47563  
A:Experimental source: strain 22/58 'Valente'  
C:Genetics:  
A:Gene: fcrV  
C:Superfamily: M5 protein  
C:Keywords: duplication; immunoglobulin receptor  
F:1-41/Domain: signal sequence #status predicted <SIG>  
F:42-587/Product: IgG Fc receptor II #status predicted <MAT>  
F:234-268/Region: 35-residue repeat A  
F:269-303/Region: 35-residue repeat A  
F:304-338/Region: 35-residue repeat A  
F:339-373/Region: 35-residue repeat A  
F:374-408/Region: 35-residue repeat B  
F:416-450/Region: 35-residue repeat B

Query Match 62.9%; Score 39; DB 2; Length 587;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAKLAKLAKLAKL 12  
||||| :|||  
Db 495 KLAQAELAKL 506

## RESULT 15

C64071  
H-transporting ATP synthase (EC 3.6.1.34) epsilon chain - Haemophilus influenzae (st  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 21-Jul-2000  
C:Accession: C64071  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630  
A;Accession: C64071  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-142 <TIGR>  
A;Cross-references: GB:U32730; GB:L42023; NID:G3212191; PIDN:AAC22136.1; PID:g1573456; T  
C;Superfamily: H<sup>+</sup>-transporting ATP synthase epsilon chain  
C;Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match 61.3%; Score 38; DB 2; Length 142;  
Best Local Similarity 72.7%; Pred. No. 50;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LAKLAKKLAKL 12  
Db 118 VAKLSRELAKL 128

Search completed: May 8, 2002, 07:19:31  
Job time: 99 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:28:07 ; Search time 35.72 seconds  
(without alignments)  
94.184 Million cell updates/sec

Title: CHIMERA\_X

Perfect score: 56

Sequence: 1 SMSIARLGXXXXXXXXXXXXX 23

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.invertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	87.5	395	2	Q47660
2	49	87.5	421	2	Q47661
3	49	87.5	450	2	Q57414
4	49	87.5	450	2	Q46730
5	49	87.5	476	2	Q57460
6	49	87.5	476	2	Q46731
7	49	87.5	476	2	Q57047
8	47	83.9	398	2	Q914M8
9	47	83.9	4215	5	Q9W332
10	46	82.1	249	2	Q84994
11	46	82.1	254	2	Q9R9B1
12	46	82.1	382	2	Q9WWT2
13	46	82.1	382	2	Q9S640
14	46	82.1	383	2	Q86045
15	46	82.1	383	2	Q2Z3A8
16	46	82.1	384	2	Q68143
17	46	82.1	387	2	Q9WWT1
18	46	82.1	387	2	Q9S639
19	46	82.1	387	2	Q9S638

20	46	82.1	387	2	Q9RQE9
21	46	82.1	388	2	Q9ZFK3
22	46	82.1	388	2	P70944
23	46	82.1	928	10	Q9M224
24	45	80.4	283	2	Q91128
25	45	80.4	308	2	Q9F4A1
26	45	80.4	334	5	Q9VYG0
27	45	80.4	338	2	Q9PCZ4
28	45	80.4	371	3	O13375
29	45	80.4	381	2	Q9EV42
30	45	80.4	384	10	Q9C524
31	45	80.4	402	2	P75810
32	45	80.4	449	12	Q65849
33	45	80.4	449	12	Q65877
34	45	80.4	453	12	Q92523
35	45	80.4	454	12	Q92521
36	45	80.4	461	3	Q9UWE0
37	45	80.4	522	5	Q9VZHA
38	45	80.4	595	2	Q9RJ38
39	45	80.4	598	5	O76414
40	45	80.4	650	12	Q911A7
41	45	80.4	650	12	Q9IZG0
42	45	80.4	687	2	Q9HW35
43	45	80.4	709	5	Q9W3W9
44	45	80.4	805	2	Q9A836
45	45	80.4	1138	2	Q9PKQ7

#### ALIGNMENTS

RESULT	1
Q47660	
ID	Q47660
AC	Q47660; PRELIMINARY; PRT; 395 AA.
DT	01-NOV-1996 (TRENBLrel. 01, Created)
DT	01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE	PROTEIN #4.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]
RP	PARTIAL SEQUENCE FROM N.A.
RX	MEDLINE=82049482; PubMed=6271452;
RA	Auerswald E.A., Ludwig G., Schaller H.;
RT	"Structural analysis of Tn5.;"
RL	Cold Spring Harb. Symp. Quant. Biol. 45:107-113(1981).
RN	[2]
RP	SEQUENCE OF 382-395 FROM N.A.
RX	MEDLINE=83106478; PubMed=6295884;
RA	Beck E., Ludwig G., Auerswald E.A., Reiss B., Schaller H.;
RT	"Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5.;"
RL	Gene 19:327-336(1982).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=83050973; PubMed=6291786;
RA	Johnson R.C., Yin J.C.P., Reznikoff W.S.;
RT	"Control of Tn5 transposition in Escherichia coli is mediated by protein from the right repeat.;"
RL	Cell 30:873-882(1982).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=82049532; PubMed=6271497;
RA	Rothstein S.J., Jorgensen R.A., Yin J.C.P., Yong-Di Z., Johnson R.C.,
RT	Reznikoff W.S.;
RL	"Genetic organization of Tn5.;"
RN	[5]
RP	Cold Spring Harb. Symp. Quant. Biol. 45:99-105(1981).
RX	SEQUENCE FROM N.A.
RX	MEDLINE=81162719; PubMed=6260374;

Q9rqe9	burkholderi
Q9zfk3	burkholderi
P70944	burkholderi
Q9m224	arabidopsis
Q91128	pseudomonas
Q9f4a1	erwinia amy
Q9vyg0	drosofila
Q9pcz4	xylella fas
O13375	saccharomyc
Q9ev42	alcaligenes
Q9c524	arabidopsis
P75810	escherichia
Q65849	barley yell
Q65877	barley yell
Q92523	barley yell
Q92521	barley yell
Q9uwe0	aspergillus
Q9vzha	drosofila
Q9rj38	streptomyce
O76414	caenorhabdi
Q911a7	barley yell
Q9izg0	barley yell
Q9hw35	pseudomonas
Q9w3w9	drosofila
Q9a836	caulobacter
Q9pkq7	chlamydia m

[4] SEQUENCE FROM N.A.  
RN MEDLINE=82049532; PubMed=6271497;  
RX Rothstein S.J., Jorgensen R.A., Yin J.C.P., Yong-Di Z., Johnson R.C.,  
RA Reznikoff W.S.;  
RA "Genetic organization of Tn5.";  
RT Cold Spring Harb. Symp. Quant. Biol. 45:99-105(1981).  
RL [5]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=81162719; PubMed=6260374;  
RX Rothstein S.J., Reznikoff W.S.;  
RA "The functional differences in the inverted repeats of Tn5 are caused  
RT by a single base pair nonhomology.;"  
RT Cell 23:191-199(1981).  
RL [6]  
RN

```

Query Match      87.58; Score 49; DB 2; Length 450;
Best Local Similarity 31.8%; Pred. No. 89;
Matches 7; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

Qy 2 MSIALRGCGXXXXXXXXXXXXX 23
Db 427 MAIALRGFGMDSKRTGASWGA 448

```

RESULT	6	
Q46731		
ID	Q46731	PRELIMINARY;
AC	Q46731;	PRT; 476 AA.
DT	01-NOV-1996	(TrEMBLrel. 01, Created)





RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector A.C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003447; AAF46505.1; -;  
 DR HSSP: P01132; 1EGF.  
 DR FLYBase: FBgn0030138; CG2996.  
 DR InterPro: IPR002123; Acyltransferase.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000859; CUB.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF-2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR Pfam: PF01553; Acyltransferase; 1.  
 DR Pfam: PF00431; CUB; 19.  
 DR Pfam: PF00008; EGF; 6.  
 DR SMART: SM00042; CUB; 19.  
 DR SMART: SM00179; EGF\_CA; 2.  
 DR SMART: SM00001; EGF\_like; 6.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE: PS01180; CUB; 29.  
 DR PROSITE: PS00022; EGF\_1; 3.  
 DR PROSITE: PS01186; EGF-2; 4.  
 DR PROSITE: PS01187; EGF\_CA; 3.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 4215 AA; 470517 MW; 53613E810BA8D4F1 CRC64;

Query Match 83.98; Score 47; DB 5; Length 4215;  
 Best Local Similarity 26.18; Pred. No. 3.8e+03;  
 Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SMSTARLGXXXXXXX 23

Db 2862 SLSVARGGSGFSPGVIASPSR 2884

RESULT 10  
 084994

ID 084994 PRELIMINARY; PRT; 249 AA.  
 AC 084994;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE FLAGELLIN (FRAGMENT).  
 OS Burkholderia pseudomallei (Pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC Burkholderia.  
 OX NCBI\_TaxID=28450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=E956;  
 RX MEDLINE=99093016; PubMed=9877189;  
 RA Winstanley C., Hales B.A., Corkill J.E., Gallagher M.J., Hart C.A.;  
 RT "Flagellin gene variation between clinical and environmental isolates  
 of Burkholderia pseudomallei contrasts with the invariance among  
 clinical isolates.";  
 RL J. Med. Microbiol. 47:689-694 (1998).  
 DR EMBL: AF030241; AAC27443.1; -;  
 DR InterPro: IPR001029; Flagellin\_C.  
 DR ProDom: PD000316; Flagellin\_C; 1.  
 FT NON\_TER 1  
 FT NON\_TER 249  
 SQ SEQUENCE 249 AA; 24915 MW; EFF4C8CA527174C3 CRC64;

Query Match 82.18; Score 46; DB 2; Length 249;  
 Best Local Similarity 26.18; Pred. No. 1.4e+02;  
 Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SMSTARLGXXXXXXX 23

Db 67 SMSAAKIGGLVQTGLGTFRV 89

RESULT 11

ID 09R9B1 PRELIMINARY; PRT; 254 AA.  
 AC 09R9B1;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE FLAGELLIN (FRAGMENT).  
 OS Burkholderia pseudomallei (Pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC Burkholderia.  
 OX NCBI\_TaxID=28450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=E505;  
 RX MEDLINE=99093016; PubMed=9877189;  
 RA Winstanley C., Hales B.A., Corkill J.E., Gallagher M.J., Hart C.A.;  
 RT "Flagellin gene variation between clinical and environmental isolates  
 of Burkholderia pseudomallei contrasts with the invariance among  
 clinical isolates.";  
 RL J. Med. Microbiol. 47:689-694 (1998).  
 DR EMBL: AF030240; AAC27442.1; -;  
 DR InterPro: IPR001029; Flagellin\_C.  
 DR ProDom: PD000316; Flagellin\_C; 1.  
 FT NON\_TER 1  
 FT NON\_TER 254  
 SQ SEQUENCE 254 AA; 25363 MW; 7EEE33271B33BA71 CRC64;

Query Match 82.18; Score 46; DB 2; Length 254;  
 Best Local Similarity 26.18; Pred. No. 1.5e+02;  
 Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SMSTARLGXXXXXXX 23

Db 67 SMSAAKIGGLVQTGLGTFRV 89





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:28:07 ; Search time 12.9 Seconds  
(without alignments)  
65.371 Million cell updates/sec

Title: CHIMERA\_X

Perfect score: 56

Sequence: 1 SMSIARLGXXXXXXX 23

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	46	82.1	498	1	MEFD_XENLA
2	45	80.4	305	1	SCRK_ECOLI
3	45	80.4	450	1	V50K_BYDVP
4	45	80.4	886	1	DSCL_MOUSE
5	45	80.4	1895	1	YLK3_CAEBL
6	44	78.6	197	1	PAAD_AERPE
7	44	78.6	221	1	YM14_ARCFU
8	44	78.6	307	1	SCRK_KLEPN
9	44	78.6	307	1	SCRK_SALTY
10	44	78.6	320	1	YDJE_BACSU
11	44	78.6	496	1	IMDH_METJA
12	44	78.6	700	1	R1R1_BACSU
13	44	78.6	796	1	YS8A_CAEBL
14	44	78.6	1315	1	CHAO_DROME
15	43	76.8	162	1	COAD_XYLFA
16	43	76.8	214	1	YPU5_RHOCA
17	43	76.8	236	1	RPC1_LAMBD
18	43	76.8	270	1	DHMA_FLAS1
19	43	76.8	282	1	MIND_CHLVU
20	43	76.8	286	1	MIND_MESVI
21	43	76.8	351	1	PERX_SOLTU
22	43	76.8	364	1	PER1_LYCES
23	43	76.8	386	1	YD46_MYCTU
24	43	76.8	392	1	YL78_ARCFU
25	43	76.8	399	1	EFTU_SYNY3
26	43	76.8	430	1	CLPX_BORHU
27	43	76.8	459	1	HN3B_MOUSE
28	43	76.8	469	1	EXOC_AZOBH
29	43	76.8	481	1	INDH_HELPJ
30	43	76.8	481	1	INDH_HELPY
31	43	76.8	509	1	SHT_LYMST
32	43	76.8	565	1	UBPN_HUMAN
33	43	76.8	566	1	UBPN_MOUSE

34	43	76.8	623	1	YD54_MYCTU	Q11024 mycobacteri
35	43	76.8	643	1	R2C1_HUMAN	P04264 homo sapien
36	43	76.8	746	1	KHTA_RHIME	Q923q5 rhizobium m
37	43	76.8	789	1	VIB4_AGR9	P05353 agrobacteri
38	43	76.8	856	1	AAP1_YEAST	P37898 saccharomyc
39	43	76.8	949	1	AHM6_ARATH	Q9szc9 arabidopsis
40	42	75.0	92	1	YTR6_AZOBR	P26944 azospirilli
41	42	75.0	193	1	DDPX_ECOLI	P77790 escherichia
42	42	75.0	264	1	FLIR_SALTY	P54702 salmonella
43	42	75.0	305	1	RCEM_RHORU	P10718 rhodospiril
44	42	75.0	307	1	QOX2_ACEAC	P50653 acetobacter
45	42	75.0	327	1	E13D_HORVU	Q02437 hordeum vul

ALIGNMENTS

RESULT 1

MEFD\_XENLA

ID MEFD\_XENLA STANDARD; PRT; 498 AA.

AC Q03413;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE MYOCYTE-SPECIFIC ENHANCER FACTOR 2D HOMOLOG (SERUM RESPONSE FACTOR-LIKE PROTEIN 1) (SL-1).

GN MEF2D OR SLL.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Neurula;

RX MEDLINE=93099873; PubMed=1281451;

RA Chambers A.E., Kotecha S., Towers N., Mohun T.J.;

RT "Muscle-specific expression of SRF-related genes in the early embryo of Xenopus laevis".

RL EMBO J. 11:4981-4991(1992).

CC -!- FUNCTION: MAY REGULATE MUSCLE-SPECIFIC TRANSCRIPTION IN THE EMBRYO AND MAY REGULATE TRANSCRIPTION OF A VARIETY OF CELL TYPES IN THE ADULT. IT BINDS TO THE SEQUENCE CTA(T/A)4TAR.

CC -!- SUBUNIT: BINDS DNA AS A MULTIMER, PROBABLY AS A DIMER.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- TISSUE SPECIFICITY: RESTRICTED TO THE SOMITIC MESODERM OF EARLY EMBRYOS AND TO THE BODY MUSCLE (MYOTOMES) OF THE TADPOLE.

CC -!- EXPRESSED IN ALL TISSUES EXAMINED IN THE ADULT.

CC -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN THE LATE GASTRULA.

CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION FACTORS. MEF2 SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; Z19124; CAA79531.1; -.

CC HSSP; P11831; 1SR5.

DR InterPro; IPR002100; MADS-box.

DR Pfam; PF00319; SRF-TF; 1.

DR PRINTS; PR00404; MADSDOMAIN.

DR SMART; SM00432; MADS; 1.

DR PROSITE; PS00350; MADS\_BOX\_1; 1.

DR PROSITE; PS0066; MADS\_BOX\_2; 1.

KW Transcription regulation; Nuclear protein; DNA-binding; Activator;

KW Multigene family.

FT DOMAIN 3 57 MADS.

FT DNA\_BIND 58 86 MEF2-TYPE (POTENTIAL).

FT DOMAIN 3 31 ARG/LYS-RICH (BASIC).



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X97986; CAA56628.1; -;  
DR EMBL; X97986; CAA56629.1; -;  
DR HSSP; P09803; 1EDH.  
DR MGD; MGI:109173; Dsc1.  
DR InterPro; IPR002126; Cadherin.  
DR Pfam; PF00028; cadherin; 5.  
DR PRINTS; PR00205; CADHERIN.  
DR SMART; SM00112; CA; 4.  
DR PROSITE; PS00232; CADHERIN\_1; 2.  
DR PROSITE; PS0268; CADHERIN\_2; 4.  
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
KW Calcium-binding; Repeat; Alternative splicing.  
FT SIGNAL 1 29 POTENTIAL.  
FT PROPEP 30 134  
FT CHAIN 135 886 DESMOCOLLIN LA/1B.  
FT DOMAIN 135 691 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 692 714 POTENTIAL.  
FT DOMAIN 715 886 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 135 242 CADHERIN 1.  
FT DOMAIN 243 354 CADHERIN 2.  
FT DOMAIN 355 471 CADHERIN 3.  
FT DOMAIN 472 575 CADHERIN 4.  
FT DOMAIN 576 682 CADHERIN 5.  
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 822 832 KVLGCGAEEH -> ESRGHTLIKN (IN ISOFORM 1B).  
FT VARSPLIC 840 886 MISSING (IN ISOFORM 1B).  
SQ SEQUENCE 886 AA; 98953 MW; F34F8D8578CE92F7 CRC64;

Query Match 80.4%; Score 45; DB 1; Length 886;  
Best Local Similarity 26.1%; Pred. No. 1.7e+02;  
Matches 6; Conservative 15; Mismatches 2; Indels 0; Gaps 0;

OY 1 SMSIARLGGXXXXXXX 23  
III: |||:|||||:|||||:  
DB 764 SMSVGTGLGGIKTQSQSFENVKG 786

RESULT 5  
ID YLK3\_CAEEL STANDARD; PRT; 1895 AA.  
AC P41951;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE D1044.3 IN CHROMOSOME III  
DE (EC 2.7.1.1-).  
GN D1044.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Pauley A., Waterston R.;  
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC STRONG, TO ZC84.1.  
CC -1- SIMILARITY: CONTAINS ? EGF-LIKE DOMAINS.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U00065; AAA50735.1; -;  
DR HSSP; Q63450; 1A06.  
DR WormPep; D1044.3; CE01206.  
DR InterPro; IPR002899; EB.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam; PF01683; EB; 12.  
DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR SMART; SM00289; WRI; 12.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS01186; EGF\_2; 2.  
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;  
KW ATP-binding; EGF-like domain.  
FT DOMAIN 431 703 PROTEIN KINASE.  
FT NP\_BIND 437 445 ATP (BY SIMILARITY).  
FT BINDING 461 461 ATP (BY SIMILARITY).  
FT ACT\_SITE 569 569 BY SIMILARITY.  
SQ SEQUENCE 1895 AA; 208393 MW; F23C9F788135AD6 CRC64;

Query Match 80.4%; Score 45; DB 1; Length 1895;  
Best Local Similarity 26.1%; Pred. No. 4.1e+02;  
Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMSIARLGGXXXXXXX 23  
I: |||:|||||:|||||:  
DB 264 SLEIAQLGGLSLIPAAISNPXP 286

RESULT 6  
ID PAAD\_AERPE STANDARD; PRT; 197 AA.  
AC QYBF0;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROBABLE AROMATIC ACID DECARBOXYLASE (EC 4.1.1.-).  
GN APE1647.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;  
OC Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
Crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).

CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /  
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC -----
DR EMBL; AP000062; BAA80648.1; -.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 197 AA; 21184 MW; BB878C6DB14C4B7F CRC64;

Query Match 78.6%; Score 44; DB 1; Length 197;
Best Local Similarity 22.7%; Pred. No. 48;
Matches 5; Conservative 17; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSIALRGXXXXXXX 23
:::|||||:
DB 142 LAIARMGIVPLTSLFYIKPS 163

RESULT 7
YMI4_ARCFU
ID YMI4_ARCFU STANDARD; PRT; 221 AA.
AC O28069;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHETICAL PROTEIN AF2214.
GN AF2214.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE000952; AAB89038.1; -.
DR TIGR; AF2214; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 41 63 POTENTIAL.
FT TRANSMEM 78 100 POTENTIAL.
FT TRANSMEM 141 163 POTENTIAL.
FT TRANSMEM 178 200 POTENTIAL.
SQ SEQUENCE 221 AA; 24925 MW; 888043795312594C CRC64;

Query Match 78.6%; Score 44; DB 1; Length 221;
Best Local Similarity 28.6%; Pred. No. 55;
Matches 6; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

QY 3-STIALRGXXXXXXX 23
:::|||||:
DB 142 LAIARMGIVPLTSLFYIKPS 163
```

```
DB 101 AIALRGVLIYLEKTINELIG 121
:::|||||:
RESULT 8
SCRK_KLEPN
ID SCRK_KLEPN STANDARD; PRT; 307 AA.
AC P26420;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE FRUCTOKINASE (EC 2.7.1.4).
DE SCRK.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1033-5P14 / KAY2026;
RX MEDLINE=92236409; PubMed=1809835;
RA Aulkemeyer P., Ebner R., Heilenmann G., Jahreis K., Schmid K.,
RA Wrieden S., Lengele J.W.;
RT "Molecular analysis of two fructokinases involved in sucrose
RT metabolism of enteric bacteria.";
RL Mol. Microbiol. 5:2913-2922(1991).
CC -!- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE = ADP + D-FRUCTOSE
CC -!- PHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X61004; CAA43322.1; -.
DR PIR; S16037; S16037.
DR PIR; S18523; S18523.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF00294; pfkb; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
KW Transferase; Kinase.
SQ SEQUENCE 307 AA; 32343 MW; F8CB20B234B262FA CRC64;

Query Match 78.6%; Score 44; DB 1; Length 307;
Best Local Similarity 21.7%; Pred. No. 80;
Matches 5; Conservative 17; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXX 23
:::|||||:
DB 35 AVGVARLGDSGFIGRVGDDPFG 57
AVGVARLGDSGFIGRVGDDPFG 57

RESULT 9
SCRK_SALTY
ID SCRK_SALTY STANDARD; PRT; 307 AA.
AC P26984;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE FRUCTOKINASE (EC 2.7.1.4).
DE SCRK.
OS Salmonella typhimurium.
OG Plasmid pUR400.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
```



```
RP SEQUENCE FROM N.A.
RX MEDLINE=92236409; PubMed=1809835;
RA Aulkeneyer P., Ebner R., Heilmann G., Jahreis K., Schmid K.,
RA Wrieden S., Lengeler J.W.;
RT "Molecular analysis of two fructokinases involved in sucrose
RL metabolism of enteric bacteria.";
RL Mol. Microbiol. 5:2913-2922(1991).
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE = ADP + D-FRUCTOSE
CC 6-PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61005; CA443323.1; -.
DR PIR; S16044; S16044.
DR PIR; S18524; S18524.
DR InterPro; IPR002173; pfkB.
DR Pfam; PF00294; pfkB; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
KW Transferase; Kinase; Plasmid.
SQ SEQUENCE 307 AA; 32916 MW; E01CB770CE20B329 CRC64;

Query Match 78.6%; Score 44; DB 1; Length 307;
Best Local Similarity 21.7%; Pred. No. 80;
Matches 5; Conservative 17; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXXXXXXXXX 23
Db 35 AVGVARLGGSGFGICAVGGDPFG 57

RESULT 10
YDJE_BACSU STANDARD; PRT; 320 AA.
AC 034768;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL SUGAR KINASE IN GUTA-COTA INTERGENIC REGION.
GN YDJE OR FRUC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / MARBURG;
RX MEDLINE=98116660; PubMed=9455482;
RA Kasahara Y., Nakai S., Ogasawara N., Yata K., Sadaie Y.;
RT "Sequence analysis of the groEL-cota region of the Bacillus subtilis
RL genome, containing the restriction/modification system genes.";
RL DNA Res. 4:335-339(1997).
CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB007638; BAA22760.1; -.
DR EMBL; Z99107; CAB12436.1; -.
DR Subtilist; BG12796; ydJE.
```

```
DR InterPro; IPR002173; pfkB.
DR Pfam; PF00294; pfkB; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
KW Hypothetical protein; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 320 AA; 34256 MW; A54E095039953B7A CRC64;

Query Match 78.6%; Score 44; DB 1; Length 320;
Best Local Similarity 26.1%; Pred. No. 83;
Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXXXXXXXXX 23
Db 43 SAAATKLGDAAFSGKVGKDPFG 65

RESULT 11
IMDH_METJA STANDARD; PRT; 496 AA.
AC Q59011;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB OR MJ1616
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: IMP IS THE RATE LIMITING ENZYME IN THE DE NOVO SYNTHESIS
CC OF GUANINE NUCLEOTIDES AND THEREFORE IS INVOLVED IN THE REGULATION
CC OF CELL GROWTH (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67602; RAB99638.1; -.
DR HSPSP; P50097; IAK5.
DR TIGR; MJ1616; -.
DR InterPro; IPR000644; CBS.
DR InterPro; IPR003009; FMN_enzyme.
DR InterPro; IPR001093; IMP_DH_GMP_RED.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF00478; IMPDH_C; 1.
DR Pfam; PF01574; IMPDH_N; 1.
```

```
DR SMART; SM00116; CBS; 2.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
KW Oxidoreductase; NAD; GMP biosynthesis; Purine biosynthesis; Repeat;
KW CBS domain; Complete proteome.
FT DOMAIN 43 95 CBS 1.
FT DOMAIN 103 155 CBS 2.
FT BINDING 306 306 IMP (POTENTIAL).
SQ SEQUENCE 496 AA; 53316 MW; C3E03FDDF3898396 CRC64;

Query Match 78.6%; Score 44; DB 1; Length 496;
Best Local Similarity 21.7%; Pred. No. 1.4e+02;
Matches 5; Conservative 18; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXX 23
Db 61 AIALRGLGLGVIRHNSIEQV 83

RESULT 12
RIRL_BACSU STANDARD; PRT; 700 AA.
AC P50620;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE).
GN NRDE OR NRDA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124182; PubMed=8969495;
RA Scotti C.; Valbuza A.; Perego M.; Galizzi A.; Albertini A.M.;
RT "The Bacillus subtilis genes for ribonucleotide reductase are similar
RT to the genes for the second class I NrdE/NrdF enzymes of
RT Enterobacteriaceae."
RL Microbiology 142:2995-3004(1996).
CC -!- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY
CC FOR DNA SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC LARGE CHAIN FAMILY. MORE SIMILAR TO ENTEROBACTERIAL NRDE THAN TO
CC NRDA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z68500; CAA92810.1; -
CC EMBL; Z99113; CAB13622.1; -
CC Subtilist; BG11404; nrde.
CC InterPro; IPR000788; Ribonucleo_red.
CC Pfam; PF00317; ribonucleo_red; 1.
DR PROSITE; PS00089; RIBORED_LARGE; 1.
KW Oxidoreductase; DNA replication; Complete proteome.
SQ SEQUENCE 700 AA; 80688 MW; D2D4B914B97BBFA6 CRC64;
```

```
Query Match 78.6%; Score 44; DB 1; Length 700;
Best Local Similarity 21.7%; Pred. No. 2e+02;
Matches 5; Conservative 17; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXX 23
Db 189 SMOLSKLGGVSLNLSKRAKE 211

RESULT 13
YSBA_CAEEL STANDARD; PRT; 796 AA.
AC Q09625;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 84.3 KDA PROTEIN ZK945.10 IN CHROMOSOME II.
GN ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48582; CAA88469.1; -
CC EMBL; Z48544; CAA88469.1; JOINED.
CC EMBL; Z48544; CAA88444.1; -
CC EMBL; Z48582; CAA88444.1; JOINED.
CC WormPep; ZK945.10; CE01732.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 30 POTENTIAL.
FT DOMAIN 273 546 SER/THR-RICH.
FT DOMAIN 656 752 SER/THR-RICH.
SQ SEQUENCE 796 AA; 84306 MW; 76DC5B03EG357A6A CRC64;

Query Match 78.6%; Score 44; DB 1; Length 796;
Best Local Similarity 26.1%; Pred. No. 2.3e+02;
Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXX 23
Db 609 SVSMPRLGGTYFSTVPGNYT 631

RESULT 14
CHAO_DROME STANDARD; PRT; 1315 AA.
AC P12024; Q9VA01;
DT 01-OCT-1989 (Rel. 12, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHAOPTIN PRECURSOR (PHOTORECEPTOR CELL-SPECIFIC MEMBRANE PROTEIN).
GN CHP OR CHT OR CG1744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```



```

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- FUNCTION: REVERSIBLY TRANSFERS AN ADENYLYL GROUP FROM ATP TO 4'-
CC -1- PHOSPHOPANTHETHEINE, YIELDING DEPHOSPHO-COA (DPCOA) AND
CC PYROPHOSPHATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + PANTHETHEINE 4'-PHOSPHATE = DIPHOSPHATE +
CC DEPHOSPHO-COA.
CC -1- PATHWAY: COENZYME A (COA) BIOSYNTHESIS; FOURTH STEP.
CC -1- SUBUNIT: HOMOHXAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COAD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003936; AAF83790.1; -.
DR InterPro; IPR001980; LPS_biosynth.
DR InterPro; IPR001994; Cytidylyltransf.
DR Pfam; PF01467; Cytidylyltransf; 1.
DR PRINTS; PR01020; LPSBIOSNTHSS.
KW transferase; Nucleotidyltransferase; Coenzyme A biosynthesis;
KW Complete proteome.
SQ SEQUENCE 162 AA; 17568 MW; 519E4052D4E364FD CRC64;

Query Match 76.8%; Score 43; DB 1; Length 162;
Best Local Similarity 30.0%; Pred. NO. 58;
Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 4 IARLGXXXXXXXXXXXXXX 23
Db 139 IARLGDVSGFAPAAVVAL 158
|||||:.....:
Search completed: May 8, 2002, 07:30:47
Job time: 160 sec

```

---

**THIS PAGE BLANK (USPTO)**



QY 2 MSIALRGXXXXXXX 23  
 II: |||||:|||||:  
 Db 292 MSVRLGGVQAHSLTPVWS 313

## RESULT 3

T47846

Arm repeat containing protein-like - Arabidopsis thaliana  
 N:Alternate names: protein T8B10.10  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T47846  
 submitted to the Protein Sequence Database, March 2000  
 A:Reference number: 224478  
 A:Accession: T47846  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-928 <RIE>  
 A:Cross-references: EMBL:AL138646  
 A:Experimental source: Cultivar Columbia; BAC clone T8B10  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 504/3; 531/3; 575/3; 647/3; 679/2; 700/3; 759/2; 861/3  
 A:Note: T8B10.10

Query Match 82.1%; Score 46; DB 2; Length 928;  
 Best Local Similarity 26.1%; Pred. No. 2e+02;  
 Matches 6; Conservative 15; Mismatches 2; Indels 0; Gaps 0;

QY 1 SMSIALRGXXXXXXX 23  
 II: |||||:|||||:  
 Db 548 SMEVARAGGVHVLMLRNCKYE 570

## RESULT 4

C83338

hypothetical protein PA2454 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 13-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: C83338  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Loy, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: C83338  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-283 <STO>  
 A:Cross-references: GB:AE004673; GB:AE004091; NID:g9948501; PIDN:AAG05842.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA2454

Query Match 80.4%; Score 45; DB 2; Length 283;  
 Best Local Similarity 26.1%; Pred. No. 84;  
 Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIALRGXXXXXXX 23  
 II: |||||:|||||:  
 Db 65 SLSLARLGARVGVGDIGSAIDF 87

## RESULT 5

S52161

probable fructokinase (EC 2.7.1.4) - Escherichia coli  
 C:Species: Escherichia coli  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S52161  
 R:Bockmann, J.  
 submitted to the EMBL Data Library, September 1994  
 A:Reference number: S52160  
 A:Accession: S52161  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-305 <BOC>  
 A:Cross-references: EMBL:X81461; NID:g608705; PIDN:CAA57218.1; PID:g608707  
 C:Superfamily: ribokinase  
 C:Keywords: phosphotransferase

Query Match 80.4%; Score 45; DB 1; Length 305;  
 Best Local Similarity 26.1%; Pred. No. 91;  
 Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIALRGXXXXXXX 23  
 II: |||||:|||||:  
 Db 35 AVGIARLGTSFGIGVGGDPFG 57

## RESULT 6

B82660

fructokinase XF1610 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: B82660  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82660  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-338 <SIM>  
 A:Cross-references: GB:AE003988; GB:AE003849; NID:g9106653; PIDN:AAF84419.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000.

Query Match 80.4%; Score 45; DB 2; Length 338;  
 Best Local Similarity 21.7%; Pred. No. 1e+02;  
 Matches 5; Conservative 18; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIALRGXXXXXXX 23  
 II: |||||:|||||:  
 Db 51 AVAVARLGGAOVGVGLGRDMFG 73

## RESULT 7

G96689

probable fructokinase F28C11.11 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: G96689



Query Match 80.4%; Score 45; DB 2; Length 402;  
Best Local Similarity 26.1%; Pred. No. 1.2e+02;  
Matches 6; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

N;Alternate names: ORF5 protein  
C;Species: barley yellow dwarf virus, BYDV  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 20-Jun-2000  
C;Accession: JQ1419  
R;Ueng, P.P.; Vincent, J.R.; Kawata, E.E.; Lei, C.H.; Lister, R.M.; Larkins, B.A.  
J. Gen. Virol. 73, 487-492, 1992  
A;Title: Nucleotide sequence analysis of the genomes of the MAV-PS1 and P-PAV isolates  
A;Reference number: JQ1409; MUID:92166764

Query Match 80.4%; Score 45; DB 2; Length 449;  
Best Local Similarity 21.7%; Pred. No. 1.4e+02;  
Matches 5: Conservative 18; Mismatches 0; Indels

A;Note: 422-lie was also round  
C;Superfamily: potato leaf roll virus 80K protein; potato leaf roll virus coat protein



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:28:02 ; Search time 19.55 seconds  
(without alignments)  
26.474 Million cell updates/sec

Title: CHIMERA\_X  
Perfect score: 56  
Sequence: 1 SMSIARLGXXXXXXX 23

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	87.5	476	2	US-08-850-880-2
2	49	87.5	476	2	US-08-944-916-2
3	49	87.5	476	2	US-08-814-877-2
4	43	76.8	236	1	US-08-333-901-2
5	43	76.8	236	1	US-08-456-582-2
6	43	76.8	236	2	US-08-898-789-2
7	43	76.8	271	2	US-07-637-865-2
8	42	75.0	3170	3	US-08-439-009A-5
9	42	75.0	3170	3	US-08-826-611-6
10	41	73.2	328	3	US-08-309-512-11
11	41	73.2	493	5	PCT-US92-08756A-11
12	41	73.2	493	5	US-08-457-274A-22
13	41	73.2	507	1	US-08-457-274A-22
14	41	73.2	507	5	PCT-US95-05758-22
15	41	73.2	545	4	US-09-066-046-39
16	41	73.2	554	4	US-08-180-371-2
17	41	73.2	626	2	US-08-596-300A-7
18	41	73.2	626	2	US-08-596-300A-14
19	41	73.2	1220	2	US-08-680-326-38
20	40	71.4	185	4	US-09-153-586-2
21	40	71.4	247	3	US-09-167-434-3
22	40	71.4	247	3	US-08-853-755-3
23	40	71.4	292	3	US-09-320-878-14
24	40	71.4	292	4	US-09-105-537-12
25	40	71.4	306	1	US-08-279-754-2
26	40	71.4	306	5	PCT-US95-09052-2
27	40	71.4	350	1	US-08-458-067-2

28	40	71.4	350	5	PCT-US96-07795-2	Sequence 2, Appli
29	40	71.4	350	5	PCT-US96-07796-2	Sequence 2, Appli
30	40	71.4	432	4	US-09-306-595C-8	Sequence 8, Appli
31	40	71.4	635	1	US-07-832-855-2	Sequence 2, Appli
32	40	71.4	660	4	US-09-181-706-8	Sequence 8, Appli
33	40	71.4	660	4	US-09-458-791-8	Sequence 8, Appli
34	40	71.4	660	4	US-09-459-066-8	Sequence 8, Appli
35	40	71.4	703	4	US-08-910-925-4	Sequence 4, Appli
36	40	71.4	717	4	US-08-910-925-1	Sequence 1, Appli
37	40	71.4	743	4	US-08-910-925-3	Sequence 3, Appli
38	40	71.4	755	4	US-09-097-767A-38	Sequence 38, Appli
39	40	71.4	775	1	US-07-603-133B-13	Sequence 13, Appli
40	40	71.4	775	1	US-07-603-133B-15	Sequence 15, Appli
41	40	71.4	900	2	US-08-813-940-4	Sequence 4, Appli
42	40	71.4	1007	2	US-08-551-459-4	Sequence 4, Appli
43	40	71.4	2101	1	US-08-466-390-4	Sequence 4, Appli
44	40	71.4	2101	1	US-08-470-950-4	Sequence 4, Appli
45	40	71.4	2101	1	US-08-467-781-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-850-880-2  
; Sequence 2, Application US/08850880  
; Patent No. 5925545  
; GENERAL INFORMATION:  
; APPLICANT: Reznikoff, William S  
; APPLICANT: Gorysin, Igor Y  
; APPLICANT: Zhou, Hong  
; TITLE OF INVENTION: System for In Vitro Transposition  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/850,880  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berson, Bennett J  
; REGISTRATION NUMBER: 37094  
; REFERENCE/DOCKET NUMBER: 960296.94142  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608/251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 476 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-850-880-2

Query Match 87.5%; Score 49; DB 2; Length 476;  
Best Local Similarity 31.8%; Pred. No. 57;  
Matches 7; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

Qy 2 MSIAIRLGXXXXXXX 23  
I:|||||:|||||:  
Db 427 MAIARLGFMDSKRTGIASWA 448

RESULT 3  
US-08-814-877-2  
; Sequence 2, Application US/08814877  
; Patent No. 5965443  
; GENERAL INFORMATION:  
; APPLICANT: Goryshin, Igor Y  
; APPLICANT: Reznikoff, William S  
; TITLE OF INVENTION: System for In Vitro Transposition  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI

Query Match 87.5%; Score 49; DB 2; Length 476;  
Best Local Similarity 31.8%; Pred. No. 57;  
Matches 7; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-333-901-2

Query Match          76.8%; Score 43; DB 1; Length 236;
Best Local Similarity 30.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY  4 IARLGGXXXXXXXXXXXX 23
    |||||:|||||:|||||:
Db  182 IARLGGDEFTFKLIRDSGQ 201

RESULT 5
US-08-456-582-2
; Sequence 2, Application US/08456582
; Patent No. 5721133
; GENERAL INFORMATION:
; APPLICANT: Dasmahapatra, Bimalendu
; TITLE OF INVENTION: No. 5721133el Protease Assays
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,789
; FILING DATE: 23-JUL-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/923,988
; FILING DATE: 21-SEP-1992
; APPLICATION NUMBER: PCT/US91/02283
; FILING DATE: 10-APR-1991
; APPLICATION NUMBER: 07/509,007
; FILING DATE: 13-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Dulak, No. 5891635man C.
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-898-789-2

Query Match          76.8%; Score 43; DB 2; Length 236;
Best Local Similarity 30.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY  4 IARLGGXXXXXXXXXXXX 23
    |||||:|||||:|||||:
Db  182 IARLGGDEFTFKLIRDSGQ 201

RESULT 7
US-07-637-865-2
; Sequence 2, Application US/07637865
; Patent No. 5942427
; GENERAL INFORMATION:
; APPLICANT: Otake, Hideko
; APPLICANT: Koyama, Yasuji
; APPLICANT: Horiuchi, Tatsuo
; APPLICANT: Nakano, Eiichi
; TITLE OF INVENTION: N-Acetylmannosamine Dehydrogenase Gene and
; TITLE OF INVENTION: No. 5942427el Recombinant DNA as Well as a Method for
; TITLE OF INVENTION: Production of N-Acetylmannosamine Dehydrogenase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: One Thomas Circle, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/637,865
; FILING DATE: 19901227
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: JP 01-338267  
FILING DATE: 28-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Skerpon, Joseph M.  
REGISTRATION NUMBER: 29,864  
REFERENCE/DOCKET NUMBER: 0118.033382  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-296-5500  
TELEFAX: 202-296-7830  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 271 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-637-865-2

Query Match 76.8%; Score 43; DB 2; Length 271;  
Best Local Similarity 23.8%; Pred. No. 2.4e+02;  
Matches 5; Conservative 16; Mismatches 0; Indels 0; Gaps 0;

QY 3 STARLGXXXXXXXXXXXX 23  
Db 81 AVARIGGLDILVAGALKGGT 101

RESULT 8  
US-07-642-734C-5  
Sequence 5, Application US/07642734C  
Patent No. 5824513  
GENERAL INFORMATION:  
APPLICANT: Katz, L  
APPLICANT: Donadio, S  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
TITLE OF INVENTION: Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Edward H. Gorznan  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckers, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-642-734C-5

Query Match 75.0%; Score 42; DB 2; Length 3170;  
Best Local Similarity 26.1%; Pred. No. 1.6e+04;

Matches 6; Conservative 15; Mismatches 2; Indels 0; Gaps 0;  
QY 1 SMSIARLGXXXXXXXXXXXX 23  
Db 668 SRVLARLGQGGMASFGLGTEQA 690  
RESULT 9  
US-08-439-009A-5  
Sequence 5, Application US/08439009A  
Patent No. 6004787  
GENERAL INFORMATION:  
APPLICANT: Donadio, S  
APPLICANT: Katz, L  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Method of Directing Biosynthesis of  
TITLE OF INVENTION: Specific Polyketides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steven F. Weinstock  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,009A  
FILING DATE: 11-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Casuto, Dianne  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3170 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-439-009A-5

Query Match 75.0%; Score 42; DB 3; Length 3170;  
Best Local Similarity 26.1%; Pred. No. 1.6e+04;  
Matches 6; Conservative 15; Mismatches 2; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXXXXXXX 23  
Db 668 SRVLARLGQGGMASFGLGTEQA 690

RESULT 10  
US-08-826-611-6  
Sequence 6, Application US/08826611  
Patent No. 6031154  
GENERAL INFORMATION:  
APPLICANT: Bennett, Alan B.  
APPLICANT: Kanayama, Yoshinori  
TITLE OF INVENTION: Fructokinase Genes and Their Use in  
TITLE OF INVENTION: Metabolic Engineering of Fruit Sweetness  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco



; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,611  
; FILING DATE: 05-APR-1997  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-077400US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 328 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-826-611-6

Query Match 73.2%; Score 41; DB 3; Length 328;  
Best Local Similarity 17.4%; Pred. No. 7.1e+02;  
Matches 4; Conservative 18; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARLGGXXXXXXX 23  
Db 49 AIAVTRLGGKSAFVGLGDDDFG 71

RESULT 11  
US-08-309-512-11  
; Sequence 11, Application US/08309512  
; Patent No. 5759828  
; GENERAL INFORMATION:  
; APPLICANT: Tal, Ronny  
; APPLICANT: Benziman, Moshe  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Ben-Bassat, Arie  
; APPLICANT: Calhoon, Roger D.  
; APPLICANT: Wong, Hing C.  
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 2730 Sand Hill Road  
; CITY: Menlo Park  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/309,512  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/800,218  
; FILING DATE: 29-NOV-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bortner, Scott R.  
; REGISTRATION NUMBER: 34,298  
; REFERENCE/DOCKET NUMBER: WEYR 20050 USA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-433-4150  
; TELEFAX: 415-433-8716  
; TELEX: 278356  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 493 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 854-3660  
; TELEFAX: (415) 854-3694  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 493 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Acetobacter xylinum  
; US-08-309-512-11

Query Match 73.2%; Score 41; DB 1; Length 493;  
Best Local Similarity 23.8%; Pred. No. 1.1e+03;  
Matches 5; Conservative 16; Mismatches 0; Indels 0; Gaps 0;

QY 3 SIARLGGXXXXXXX 23  
Db 217 AISRLGGDEFAILRLSLKLM 237

RESULT 12  
PCT-US92-08756A-11  
; Sequence 11, Application PC/TUS9208756A  
; GENERAL INFORMATION:  
; APPLICANT: Tal, Ronny  
; APPLICANT: Benziman, Moshe  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Ben-Bassat, Arie  
; APPLICANT: Calhoon, Roger D.  
; APPLICANT: Wong, Hing C.  
; TITLE OF INVENTION: CYCLIC DIGUANYLATE METABOLIC ENZYMES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Limbach and Limbach  
; STREET: 2001 Ferry Building  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0,  
; SOFTWARE: Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/08756A  
; FILING DATE: 19921014  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/800,218  
; FILING DATE: 29-NOV-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bortner, Scott R.  
; REGISTRATION NUMBER: 34,298  
; REFERENCE/DOCKET NUMBER: WEYR 20050 USA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-433-4150  
; TELEFAX: 415-433-8716  
; TELEX: 278356  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 493 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Acetobacter xylinum  
PCT-US92-08756A-11

Query Match 73.2%; Score 41; DB 5; Length 493;  
Best Local Similarity 23.8%; Pred. No. 1.4e+03;  
Matches 5; Conservative 16; Mismatches 0; Indels 0; Gaps 0;

QY 3 STARLGXXXXXXX 23  
Db 217 AISRLGDEFILRLSLKLM 237

RESULT 13  
US-08-457-274A-22  
; Sequence 22, Application US/08457274A  
; Patent No. 5734086  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Jeffrey G.  
; APPLICANT: Tomita, Takashi  
; TITLE OF INVENTION: Cytochrome P450lpr Gene and Its Uses  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,274A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/240 (D-1519)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1304  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Musca domestica  
; STRAIN: Rutgers  
; DEVELOPMENTAL STAGE: Adult  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: Chromosome 5  
; US-08-457-274A-22

Query Match 73.2%; Score 41; DB 1; Length 507;  
Best Local Similarity 23.8%; Pred. No. 1.4e+03;  
Matches 5; Conservative 16; Mismatches 0; Indels 0; Gaps 0;

QY 3 STARLGXXXXXXX 23  
Db 283 SISKLGGLTFNELAQVFFV 303

RESULT 14  
PCT-US95-05758-22  
; Sequence 22, Application PC/TUS9505758  
; GENERAL INFORMATION:  
; APPLICANT: Cornell Research Foundation, Inc.  
; TITLE OF INVENTION: Cytochrome P450lpr Gene and Its  
; TITLE OF INVENTION: Uses  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/05758  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/241 (D-1519)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-263-1304  
; TELEFAX: 716-263-1600  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 507 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Musca domestica  
; STRAIN: Rutgers  
; DEVELOPMENTAL STAGE: Adult  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: Chromosome 5  
; PCT-US95-05758-22

Query Match 73.2%; Score 41; DB 5; Length 507;  
Best Local Similarity 23.8%; Pred. No. 1.4e+03;  
Matches 5; Conservative 16; Mismatches 0; Indels 0; Gaps 0;

QY 3 STARLGXXXXXXX 23  
Db 283 SISKLGGLTFNELAQVFFV 303

RESULT 15  
US-09-066-046-39  
; Sequence 39, Application US/09066046A  
; Patent No. 6204252  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
; EHRLLICHIA AND METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:

Search completed: May 8, 2002, 07:29:16  
Job time: 74 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:28:02 ; Search time 38.64 seconds  
(without alignments)  
44.091 Million cell updates/sec

Title: CHIMERA\_X

Perfect score: 56

Sequence: 1 SMSIARLGCGXXXXXXXXXXXXX 23

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_ll01:\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	23	21 AAB21938	Prostate homing an
2	56	100.0	23	22 AAE06513	Chimeric prostate-
3	49	87.5	49	11 AAR07414	pADB-neo-ARE trans
4	49	87.5	476	19 AA056694	Modified (transpos
5	49	87.5	476	20 AAY42535	Mutant Tn5 transpo
6	49	87.5	476	20 AAY42536	Wildtype Tn5 trans
7	49	87.5	476	20 AAY42537	Mutant Tn5 transpo
8	49	87.5	476	20 AAY42538	Mutant Tn5 transpo
9	49	87.5	476	20 AAY42539	Mutant Tn5 transpo
10	49	87.5	476	20 AAY42540	Mutant Tn5 transpo
11	49	87.5	476	20 AAY42541	Mutant Tn5 transpo

12	49	87.5	476	20	AAV42542	Mutant Tn5 transpo
13	49	87.5	476	20	AAV42543	Mutant Tn5 transpo
14	49	87.5	476	20	AAV42544	Mutant Tn5 transpo
15	49	87.5	476	20	AAV15382	Modified Tn5 trans
16	49	87.5	476	22	AA61988	Wild-type Tn5 tran
17	45	80.4	142	21	AA32837	Eucalyptus grandis
18	45	80.4	382	21	AA82494	BYDV coat protein
19	45	80.4	389	22	AA878883	C. glutamicum SRT
20	45	80.4	484	22	AA92356	C. glutamicum prote
21	45	80.4	484	22	AA878882	C. glutamicum SRT
22	45	80.4	651	21	AA82492	BYDV coat protein
23	45	80.4	724	21	AA82493	BYDV coat protein
24	45	80.4	1416	20	AA00211	Enterococcus faeca
25	45	80.4	1448	20	AA00210	Enterococcus faeca
26	44	78.6	109	22	AA22026	Peptide #8460 enco
27	44	78.6	109	22	AA38392	Peptide #12429 enc
28	44	78.6	166	21	AA825307	Eucalyptus grandis
29	44	78.6	268	22	AA89888	C glutamicum prote
30	44	78.6	314	22	AA93055	C glutamicum prote
31	44	78.6	1622	21	AA10483	S. aggregatum PKS
32	44	78.6	2870	21	AA95559	Caenorhabditis ele
33	44	78.6	3178	21	AA95556	Caenorhabditis ele
34	43	76.8	133	21	AA24404	Arabidopsis thalia
35	43	76.8	162	21	AA25901	Zea mays protein f
36	43	76.8	231	12	AA12878	N-acetylmannosamin
37	43	76.8	236	19	AAW47125	Phage lambda repre
38	43	76.8	236	20	AAV41286	Peptide encoded by
39	43	76.8	237	19	AAW80738	Amino acid sequenc
40	43	76.8	237	19	AAW58897	Plasmid PCSJ bacte
41	43	76.8	237	19	AAW58895	Plasmid pAW12/10 b
42	43	76.8	247	21	AA30797	Arabidopsis thalia
43	43	76.8	248	21	AA30796	Arabidopsis thalia
44	43	76.8	252	21	AA39232	Arabidopsis thalia
45	43	76.8	253	21	AA39231	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AAB21938 AAB21938 standard; Peptide: 23 AA.

XX AC AAB21938;

XX DT 22-MAR-2001 (first entry)

XX DE Prostate homing antimicrobial pro-apoptotic conjugate.

XX KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;  
alpha-helix; human.

XX OS Chimeric - Homo sapiens.

XX OS Chimeric - Unidentified.

XX PH Key Location/Qualifiers

XX FT Misc-difference 10..23 /note= "Preferably D-form residues"

XX PN WO200042973-A2.

XX PD 27-JUL-2000.

XX PF 21-JAN-2000; 2000WO-US01602.

XX PR 22-JAN-1999; 99US-0235902.

XX PA (BURN-) BURNHAM INST.

XX PI Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;

XX DR WPI; 2000-499174/44.

XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
 PT selectively homes to a mammalian cell type or tissue linked to an  
 PT antimicrobial peptide, useful for the treatment of prostate cancer -  
 XX  
 PS Claim 29; Page 108; 118pp; English.  
 CC  
 CC The present invention relates to homing pro-apoptotic conjugates,  
 CC comprising of a tumor homing molecule that selectively homes to a  
 CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
 CC homing pro-apoptotic conjugates are selectively internalised by the  
 CC mammalian cell type or tissue and exhibits high toxicity, especially to  
 CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
 CC toxicity when not linked to the tumor homing molecule. In addition, the  
 CC antimicrobial peptide has an amphipathic alpha-helical structure. The  
 CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,  
 CC breast and prostate cancer or melanoma. The present sequence is one such  
 CC prostate homing pro-apoptotic conjugate.  
 XX  
 SQ Sequence 23 AA;

Query Match 100.0%; Score 56; DB 21; Length 23;  
 Best Local Similarity 39.1%; Pred. No. 0.13;  
 Matches 9; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXX 23  
 |||||  
 Db 1 smsiarlggklaklaklak 23

RESULT 2  
 AAE06513  
 ID AAE06513 standard; peptide; 23 AA.

XX AAE06513;

XX 25-SEP-2001 (first entry)

XX Chimeric prostate-homing pro-apoptotic peptide.

XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
 KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
 KW cytostatic.

XX Unidentified.

XX Key Location/Qualifiers  
 FH Domain 1..7  
 FT Domain /label= prostate\_homing\_domain  
 FT Domain 8..9  
 FT /label= Coupling\_domain  
 FT /note= "Glycylglycine bridge"  
 FT Domain 10..23  
 FT /label= Antimicrobial\_peptide

XX WO200153342-A1.

XX 26-JUL-2001.

XX 16-JAN-2001; 2001WO-US01362.

XX 21-JAN-2000; 2000US-0489582.

XX (BURN-) BURNHAM INST.

XX Ruoslahti EI, Pasqualini R, Arap W, Breidesen DE, Ellerby HM;

XX WPI; 2001-451901/48.

XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
 PT prostate cancer, comprises a prostate-homing peptide linked to an  
 PT antimicrobial peptide -

XX Claim 6; Page 103; 176pp; English.  
 PS  
 CC The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used  
 CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is a chimeric  
 CC prostate-homing pro-apoptotic peptide.  
 XX  
 SQ Sequence 23 AA;

Query Match 100.0%; Score 56; DB 22; Length 23;  
 Best Local Similarity 39.1%; Pred. No. 0.13;  
 Matches 9; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARLGXXXXXXX 23  
 |||||  
 Db 1 smsiarlggklaklaklak 23

RESULT 3  
 AAR07414  
 ID AAR07414 standard; protein; 49 AA.

XX AAR07414;

XX 28-JAN-1991 (first entry)

XX pJDB-neo-ATE translation product 2.

XX Vector; regulatory sequence; plasmids; expression; ATG codon;  
 KW promoter; Neo structural gene.

XX JP02222689-A.

XX 05-SEP-1990.

XX 23-FEB-1989; 89JP-0041604.

XX 23-FEB-1989; 89JP-0041604.

XX (TOFU ) TONEN CORP.

XX WPI; 1990-315505/42.

XX N-PSDB; AAQ06275.

XX Vector for detection of regulatory sequence - by using  
 PT combination of plasmids with different expression powers etc.

XX Disclosure; Fig 1; 16pp; Japanese.

XX The DNA sequence encoding this sequence (together with 4 others,  
 CC see AAR07413 and AAR08165-67) contains the Neo structural gene or its  
 CC deriv. with, except for the region upstream of the translation  
 CC initiation codon, and the translation initiation codon itself, 7  
 CC additional ATG sequences present. A plasmid contg. this sequence can  
 CC also comprise a promoter. The combination of plasmids having different  
 CC expression powers, results in various kinds of regulatory sequences  
 CC having different expression capacities which can be measured  
 CC quantitatively. See tag b of AAQ06275.

XX Sequence 49 AA;

Query Match 87.5%; Score 49; DB 11; Length 49;  
 Best Local Similarity 31.8%; Pred. No. 5.6;  
 Matches 7; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

This sequence represents mutant Tn5 transposase EK54 MA56 LP372. This mutant form differs from the wildtype in that it binds to repeat sequences of donor DNA with greater avidity than the wildtype Tn5 transposase and it is less likely than the wildtype protein to assume an inactive multimeric structure. The improved avidity of the modified transposase for the repeat sequences for the Outside End (OE) termini is due to the Lys residue at amino acid 54, which is Glu in wild type Tn5 transposase (AAV42536). The mutation strongly alters the preference of the transposase for the OE termini, as opposed to the Inside End (IE) termini. The higher binding of this mutation, known as EK54, to OE termini results in a transposition rate that is about 10-fold higher than seen with wild-type transposase. Similar changes at position 34 to Val and position 47 (Thr to Pro change) also result in increased binding to the OE termini. The reduction in the formation of the inactive multimeric form is achieved by modifying amino acid 372 (Leu) of wildtype transposase to Pro (LP372). This sequence also has a substitution of Ala for wildtype Met at position 56. This substitution prevents the translation of an inhibitor protein, the initiation codon of which





inactive multimeric structure. The improved avidity of the modified transposase for the repeat sequences for the Outside End (OE) termini is due to the Val residue at amino acid 54, which is Glu in wild type Tn5 transposase (AAV42536). The mutation strongly alters the preference of the transposase for the OE termini, as opposed to the Inside End (IE) termini. The higher binding of this mutation, known as EY54, to OE termini results in a transposition rate that is about 3-fold higher than seen with wild-type transposase. The reduction in the formation of the inactive multimeric form is achieved by modifying amino acid 372 (Leu) of wildtype Tn5 transposase to pro (LP372). The Tn5-based transposition system is more broadly applicable and utilises shorter, more well defined termini that are active on DNA of any structure when compared to Mu- and Tn10-based systems. The modified transposase enzyme is useful in a system for introducing genetic changes to nucleic acid. The method can be applied to create absolute defective mutants, to provide selective markers to target DNA, to provide portable regions of homology to a target DNA, to facilitate insertion of specialised DNA sequences into target DNA, to provide primer binding sites or tags for DNA sequencing, and to facilitate production of genetic fusions for gene expression studies and protein domain mapping.

Note: The present sequence is not shown in the specification but is derived from the mutant Tn5 transposase sequence given in columns 15-18.

Seq	Sequence	470 bp	87.5%	Score 49	DB 20	Length 476
	Query Match					

Seq	Matches	Conservative	Mismatches	Indels	Gaps
Ov	2	MSIARLGXXXXXXXXXXXXX	23		

DD 427 MATRATSGIMUSKI CYRUSWA 4750

RESULT 9

AA42539  
 ID AAY42539 standard; protein; 476 AA.  
 XX  
 AC AAY42539;

Strain	20-DEC-1999 (first entry)
DA	
DT	
XX	
DE	

Transposase; modified form; wildtype; mutant; multimeric; OE termini;  
 KW IE termini; outside end termini; inside end termini;  
 KW repeat sequence; mutation; mutuin.  
 KW

AA  
OS Transposon Tn5.  
OS Synthetic.

AA  
PN  
XX  
XX  
US5965443-A.

XX 09-SEP-1996; 96US-0814877.

09-SEP-1996; 9605-0814877.  
XX

XX  
PI Reznikoff WS, Goryshin IV;

DR	WPI; 1999-579916/49.
XX	
PT	Modified transposase enzyme for use in a system for introducing genetic changes to nucleic acid -
PT	

AA  
PS Disclosure; Page -; 21pp; English.  
XX

CC This sequence represents mutant M3 transposase ER54. This  
CC mutant form differs from the wildtype in that it binds to repeat







---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:18:34 ; Search time 38.66 seconds  
(without alignments)  
13.412 Million cell updates/sec

Title: US-09-765-086-207

Perfect score: 30

Sequence: 1 SMSIARL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A\_Geneseq\_1101.\*  
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	30	100.0	7	20	AAV48905	Membrane dipeptida
2	30	100.0	7	21	AAB21907	Human prostate-hom
3	30	100.0	7	22	AAE06483	Mouse prostate hom
4	30	100.0	23	21	AAB21938	Prostate homing an
5	30	100.0	23	22	AAE06513	Chimeric prostate-
6	29	96.7	236	22	AAE30583	c glutamicum prote
7	27	90.0	191	21	AAE08955	Arabidopsis thalia
8	27	90.0	289	22	AAE91999	Arabidopsis thalia
9	27	90.0	301	20	AAV37469	c glutamicum prote
10	26	86.7	183	22	AAU14179	Protein involved i
11	26	86.7	260	21	AAB41860	Human ORFX protei
						Human ORFX ORF1624

12	26	86.7	382	22	AAG92195	C glutamicum prote
13	26	86.7	382	22	AAE79265	Corynebacterium q1
14	26	86.7	422	18	AAW18032	zebrafish retinoid
15	26	86.7	697	21	AAE25565	Eucalyptus grandis
16	26	86.7	748	22	AAE40225	Human polyptide
17	25	83.3	118	22	AAG98415	Escherichia coli p
18	25	83.3	118	22	AAG98916	E. coli growth and
19	25	83.3	118	22	AAG98948	E. coli growth and
20	25	83.3	204	21	AAV75241	Neisseria gonorrhe
21	25	83.3	381	22	AAG92942	C glutamicum prote
22	24	80.0	64	21	AAE39477	Human secreted pro
23	24	80.0	120	22	AAE94277	Human protein sequ
24	24	80.0	137	18	AAW20100	H. pylori cytoplas
25	24	80.0	266	19	AAW98839	H. pylori GHPO 153
26	24	80.0	319	21	AAG04867	Arabidopsis thalia
27	24	80.0	319	21	AAG42777	Arabidopsis thalia
28	24	80.0	336	21	AAG04866	Arabidopsis thalia
29	24	80.0	336	21	AAG42776	Arabidopsis thalia
30	24	80.0	342	21	AAG04865	Arabidopsis thalia
31	24	80.0	342	21	AAG42775	Arabidopsis thalia
32	24	80.0	363	21	AAG21936	Arabidopsis thalia
33	24	80.0	381	21	AAG21935	Arabidopsis thalia
34	24	80.0	387	21	AAG21934	Arabidopsis thalia
35	24	80.0	477	15	AAE62448	Rat calcitonin rec
36	24	80.0	515	15	AAE62447	Rat calcitonin rec
37	24	80.0	2675	21	AAE07564	Human secreted pro
38	23	76.7	23	21	AAE27643	PJDB-neo-AE trans
39	23	76.7	49	11	AAE07414	Human secreted pro
40	23	76.7	62	21	AAE44338	Human secreted pro
41	23	76.7	63	21	AAE55635	Arabidopsis thalia
42	23	76.7	63	21	AAE58828	Arabidopsis thalia
43	23	76.7	81	22	AAE14873	Peptide #1307 enco
44	23	76.7	81	22	AAE27300	Peptide #1337 enco
45	23	76.7	81	22	AAE02598	Peptide #1280 enco

#### ALIGNMENTS

RESULT 1  
AAV48905  
ID AAY48905 standard; Peptide; 7 AA.

XX AC AAY48905;

XX DT 10-DEC-1999 (first entry)

XX DE Membrane dipeptidase-binding prostate homing peptide #21.

XX KW Homing peptide; organ; tissue; lung; pancreas; skin; retina; MDP;

XX KW prostate; ovary; lymph node; adrenal gland; liver; gut; tumour;

XX KW membrane dipeptidase.

XX OS Synthetic.

XX OS Homo-sapiens.

XX PN WO9946284-A2

XX PD 16-SEP-1999

XX PF 10-MAR-1999; 99WO-US05284.

XX PR 13-MAR-1998; 98US-0042107.

XX PR 26-FEB-1999; 99US-0042107.

XX PA (BURN-) BURNHAM INST.

XX PI Rajotte D, Pasqualini R, Ruoslahti EI;

XX DR WPI; 1999-571717/48.

XX PT New peptides which selectively home to organs or tissues, used for,

PT e.g. identifying target ligands and for therapy of pathological

PT conditions -  
XX Claim 1; Page 151; 193pp; English.  
XX The present invention describes peptides that selectively home to a  
XX tissue or organ. The peptides can be used for identifying an organ  
CC or tissue, for identifying a target molecule expressed by an organ or  
CC tissue or for treating an organ or tissue pathology, where the organ or  
CC tissue is selected from prostate, lung, skin, retina, pancreas, gut,  
CC ovary, adrenal gland, liver, and lymph node. The peptide bind to the  
CC membrane dipeptidase (MDP). AAY48618 to AAY49066 represent sequences  
XX which are used in the exemplification of the present invention.  
XX SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 30; DB 20; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SMSIARL 7  
Db 1 smsiarl 7  
|||||  
  
RESULT 2  
AAB21907  
ID AAB21907 standard; Peptide; 7 AA.  
XX AAB21907;  
AC AAB21907;  
XX 22-MAR-2001 (first entry)  
DT Human prostate-homing peptide #1.  
DE Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
XX breast; prostate; melanoma; cancer; Kaposi's sarcoma; human.  
KW Homo sapiens.  
OS WO200042973-A2.  
XX 27-JUL-2000.  
PD 21-JAN-2000; 2000WO-US01602.  
XX 22-JAN-1999; 99US-0235902.  
PR (BURN-) BURNHAM INST.  
XX Ellerby HM, Bredeesen DE, Pasqualini R, Ruoslahti EI;  
PI WPI; 2000-499174/44.  
XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
PT selectively homes to a mammalian cell type or tissue linked to an  
PT antimicrobial peptide, useful for the treatment of prostate cancer -  
XX Claim 25; Page 107; 118pp; English.  
XX The present invention relates to homing pro-apoptotic conjugates,  
CC comprising of a tumour homing molecule that selectively homes to a  
CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
CC homing pro-apoptotic conjugates are selectively internalised by the  
CC mammalian cell type or tissue and exhibits high toxicity, especially to  
CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
CC toxicity when not linked to the tumor homing molecule. The conjugates are  
CC useful for the treatment of cancer e.g. Kaposi's sarcoma, breast and  
CC prostate cancer or melanoma. The present sequence is a homing peptide  
CC isolated in the present invention, which can be conjugated to an  
CC antimicrobial peptide to make the homing pro-apoptotic conjugates of the  
XX present invention.

SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 30; DB 21; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SMSIARL 7  
Db 1 smsiarl 7  
|||||  
  
RESULT 3  
AAE06483  
ID AAE06483 standard; peptide; 7 AA.  
XX AAE06483;  
AC AAE06483;  
XX 25-SEP-2001 (first entry)  
DT Mouse prostate homing peptide #1.  
DE Mouse; chimeric prostate-homing pro-apoptotic peptide;  
XX prostate-homing peptide; antimicrobial peptide; prostate cancer;  
KW tumour homing molecule; cytostatic.  
XX Mus sp.  
OS WO200153342-A1.  
XX 26-JUL-2001.  
PD 16-JAN-2001; 2001WO-US01362.  
XX 21-JAN-2000; 2000US-0489582.  
PR (BURN-) BURNHAM INST.  
XX Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;  
PI WPI; 2001-451901/48.  
XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
PT prostate cancer, comprises a prostate-homing peptide linked to an  
PT antimicrobial peptide -  
XX Claim 2; Page 103; 176pp; English.  
XX The patent discloses novel chimeric prostate-homing pro-apoptotic  
CC peptide which comprises a prostate-homing peptide linked to an  
CC antimicrobial peptide, where the chimeric peptide is selectively  
CC internalised by and exhibits high toxicity to prostate tissue and  
CC where the antimicrobial peptide has low mammalian cell toxicity when  
CC not linked to prostate-homing peptide. The chimeric peptide is used  
CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
CC induce selective toxicity in vivo in a prostate cancer, and to treat  
CC a patient with prostate cancer. The present sequence is mouse prostate  
CC homing peptide. This sequence is useful in the homing of pro-apoptotic  
CC conjugates of the invention.  
XX SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 30; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SMSIARL 7  
Db 1 smsiarl 7  
|||||  
  
RESULT 4



AAB21938  
ID AAB21938 standard; Peptide; 23 AA.  
XX AC AAB21938;  
XX DT 22-MAR-2001 (first entry)  
XX DE Prostate homing antimicrobial pro-apoptotic conjugate.  
XX KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
XX KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphoteric;  
XX KW alpha-helix; human.  
XX OS Chimeric - Homo sapiens.  
XX OS Chimeric - Unidentified.  
XX FH Key Location/Qualifiers  
FT Misc-difference 10..23  
FT /note= "Preferably D-form residues"  
XX PN WO200042973-A2.  
XX PD 27-JUL-2000.  
XX PF 21-JAN-2000; 2000WO-US01602.  
XX PR 22-JAN-1999; 99US-0235902.  
XX PA (BURN-) BURNHAM INST.  
XX PI Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;  
XX WPI; 2000-499174/44.  
XX PT Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
XX PT selectively homes to a mammalian cell type or tissue linked to an  
XX PT antimicrobial peptide, useful for the treatment of prostate cancer -  
XX PS Claim 29; Page 108; 118pp; English.  
XX CC The present invention relates to homing pro-apoptotic conjugates,  
XX CC comprising of a tumour homing molecule that selectively homes to a  
XX CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
XX CC homing pro-apoptotic conjugates are selectively internalised by the  
XX CC mammalian cell type or tissue and exhibits high toxicity, especially to  
XX CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
XX CC toxicity when not linked to the tumor homing molecule. In addition, the  
XX CC antimicrobial peptide has an amphipathic alpha-helical structure. The  
XX CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,  
XX CC breast and prostate cancer or melanoma. The present sequence is one such  
XX CC prostate homing pro-apoptotic conjugate.  
XX SQ Sequence 23 AA;  
Query Match 100.0%; Score 30; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SMSIARL 7  
Db 1 smsiarl 7  
RESULT 5  
AAE06513  
ID AAE06513 standard; peptide; 23 AA.  
XX AC AAE06513;  
XX DT 25-SEP-2001 (first entry)  
XX DE Chimeric prostate-homing pro-apoptotic peptide.  
XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
XX antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
XX cytostatic.  
XX OS Unidentified.  
XX FH Key Location/Qualifiers  
FT Domain 1..7  
FT /label= Prostate\_homing\_domain  
FT Domain 8..9  
FT /label= Coupling\_domain  
FT /note= "Glycylglycine bridge"  
FT Domain 10..23  
FT /label= Antimicrobial\_peptide  
XX PN WO200153342-A1.  
XX PD 26-JUL-2001.  
XX PF 16-JAN-2001; 2001WO-US01362.  
XX PR 21-JAN-2000; 2000US-0489582.  
XX PA (BURN-) BURNHAM INST.  
XX PI Ruoslahti EI, Pasqualini R, Arap W, Bredesen DE, Ellerby HM;  
XX WPI; 2001-451901/48.  
XX PT Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
XX PT prostate cancer, comprises a prostate-homing peptide linked to an  
XX PT antimicrobial peptide -  
XX PS Claim 6; Page 103; 176pp; English.  
XX CC The patent discloses novel chimeric prostate-homing pro-apoptotic  
XX CC peptide which comprises a prostate-homing peptide linked to an  
XX CC antimicrobial peptide, where the chimeric peptide is selectively  
XX CC internalised by and exhibits high toxicity to prostate tissue and  
XX CC where the antimicrobial peptide has low mammalian cell toxicity when  
XX CC not linked to prostate-homing peptide. The chimeric peptide is used  
XX CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
XX CC induce selective toxicity in vivo in a prostate cancer, and to treat  
XX CC a patient with prostate cancer. The present sequence is a chimeric  
XX CC prostate-homing pro-apoptotic peptide.  
XX SQ Sequence 23 AA;  
Query Match 100.0%; Score 30; DB 22; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SMSIARL 7  
Db 1 smsiarl 7  
RESULT 6  
AAG90583  
ID AAG90583 standard; Protein; 236 AA.  
XX AC AAG90583;  
XX DT 26-SEP-2001 (first entry)  
XX DE C glutamicum protein fragment SEQ ID NO: 4337.  
XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX KW organic acid synthesis.  
XX OS Corynebacterium glutamicum.

XX EP1108790-A2.  
XX 20-JUN-2001.  
XX 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
XX 07-APR-2000; 2000JP-0159162.  
XX 03-AUG-2000; 2000JP-0280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI; 2001-376931/40.  
XX N-PSDB; AAH65802.  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
XX mutation point of a gene, measuring expression of a gene, analysing  
XX expression profile or pattern of a gene and identifying homologous gene  
XX  
XX Claim 17; SEQ ID NO: 4337; 246pp + Sequence Listing; English.  
XX The present invention provides a number of nucleotide and protein  
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
XX are useful for identifying the mutation point of a gene derived from a  
XX mutant of coryneform bacterium, measuring expression amount and  
XX analysing the expression profile or expression pattern of a gene derived  
XX from Coryneform bacterium, and identifying a homologue of a gene derived  
XX from coryneform bacterium. Coryneform bacteria are useful for producing  
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX particularly L-lysine. The present sequence is a protein described  
XX in the exemplification of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from the  
XX European Patent Office.  
XX Sequence 236 AA;  
XX  
XX Query Match 96.7%; Score 29; DB 22; Length 236;  
XX Best Local Similarity 85.7%; Pred. No. 13;  
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 SMSIARL 7  
XX 111111  
XX Db 2 smsvar1 8  
XX  
XX RESULT 7  
XX AAG08955  
XX ID AAG08955 standard; Protein; 191 AA.  
XX AC AAG08955;  
XX 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 6694.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125786.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 06-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
XX 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.  
XX 18-JUN-1999; 99US-0139462.  
XX 18-JUN-1999; 99US-0139463.  
XX 18-JUN-1999; 99US-0139750.  
XX 18-JUN-1999; 99US-0139763.  
XX 21-JUN-1999; 99US-0139817.  
XX 22-JUN-1999; 99US-0139899.  
XX 23-JUN-1999; 99US-0140353.  
XX 23-JUN-1999; 99US-0140354.  
XX 24-JUN-1999; 99US-0140695.  
XX 24-JUN-1999; 99US-0140823.  
XX 29-JUN-1999; 99US-0140991.  
XX 30-JUN-1999; 99US-0141287.  
XX 01-JUL-1999; 99US-0141842.  
XX 01-JUL-1999; 99US-0142154.  
XX 02-JUL-1999; 99US-0142055.  
XX 06-JUL-1999; 99US-0142390.  
XX 08-JUL-1999; 99US-0142803.  
XX 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 18-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 90.0%; Score 27; DB 21; Length 191;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
Db 12 smaiarl 18

RESULT 8  
AAG91999  
ID AAG91999 standard; Protein; 289 AA.  
XX  
AC AAG91999;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum protein fragment SEQ ID NO: 5753.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH67218.  
 XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 XX  
 PT  
 XX  
 PS Claim 17: SEQ ID NO: 5753; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 289 AA;

Query Match 90.0%; Score 27; DB 22; Length 289;  
 Best Local Similarity 85.7%; Pred. No. 50;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 DB 142 smsiarv 148  
 |||  
 |||

RESULT 9  
 AAY37469  
 ID AAY37469 standard; Protein; 301 AA.  
 XX  
 AC AAY37469;

DT 07-OCT-1999 (first entry)

DE Protein involved in intermediate metabolism of polypeptides.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

OS WO9928475-A2.

PN 10-JUN-1999.

PD 27-NOV-1998; 98WO-IB01939.

PF 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

PA Griffiths R;

XX WPI; 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis

PS Disclosure; Page 1157-1158; 1755pp; English.  
 XX  
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 SQ Sequence 301 AA;

Query Match 90.0%; Score 27; DB 20; Length 301;  
 Best Local Similarity 85.7%; Pred. No. 52;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 DB 158 smaia1 164  
 |||  
 |||

RESULT 10  
 AAU14179  
 ID AAU14179 standard; Protein; 183 AA.  
 XX  
 AC AAU14179;

DT 24-OCT-2001 (first entry)

DE Human novel protein #50.

KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
 KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;  
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.

OS Homo sapiens.

PN WO200155437-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02623.

PR 25-JAN-2000; 2000US-0491404.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

DR N-PSDB; AAS22484.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,  
 PT nervous system disorders, and for regenerating bone and cartilage -

XX Example 4; Page 555; 894pp; English.

XX The invention relates to polynucleotides encoding novel human  
 CC proteins or their active domains. The polypeptides, polynucleotides and  
 CC antibodies raised against the polypeptides are used in a method of  
 CC treatment of a mammal and prevention of disorders caused by the aberrant  
 CC protein expression or activity. The polypeptides can be used as  
 CC molecular weight markers, food supplements, and in antibody production.  
 CC The polypeptides are used to identify compounds which bind to the

CC polypeptides. Polynucleotides of the invention are used as probes and  
 CC primers, for sequencing, for chromosome or gene mapping, in the  
 CC production of recombinant proteins, and in generating anti-sense DNA or  
 CC RNA and in gene therapy. Polypeptides of the invention can be used to  
 CC target drugs to a tumour, in assays to determine biological activity, to  
 CC raise antibodies/elicit an immune response, to determine quantitative  
 CC protein levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
 CC anti-inflammatory diseases, nervous system disorders, and infection.  
 CC The present sequence represents a protein of the invention.

XX Sequence 183 AA;

Query Match 86.7%; Score 26; DB 22; Length 183;  
 Best Local Similarity 85.7%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 IIII II  
 Db 36 smsitrl 42

RESULT 11

ID AAB41860  
 XX AAB41860 standard; Protein; 260 AA.

AC AAB41860;

XX 08-FEB-2001 (first entry)

DT Human ORFX ORF1624 polypeptide sequence SEQ ID NO:3248.

DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipariatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX Homo sapiens.

OS WO200058473-A2.

PN 05-OCT-2000.

PD 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

PA Shinkets RA, Leach M;

XX

DR WPI; 2000-602362/37.  
 DR N-PSDB; AAC76069.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2449-2450; 5507pp; English.

XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipariatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 260 AA;

Query Match 86.7%; Score 26; DB 21; Length 260;  
 Best Local Similarity 71.4%; Pred. No. 80;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 IIII II  
 Db 208 smslari 214

RESULT 12

AAG92195

ID AAG92195 standard; Protein; 382 AA.

XX AAG92195;

AC AAG92195;

XX 26-SEP-2001 (first entry)

DT C glutamicum protein fragment SEQ ID NO: 5949.

DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS EP1108790-A2.

PN 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO ROGYO KK.

XX Nakagawa S, Mizoquchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX

DR WPI; 2001-376931/40.  
 XX N-PSDB; AAH67414.  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 XX  
 PS Claim 17; SEQ ID NO: 5949; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 CC Sequence 382 AA;  
 SQ

Query Match 86.7%; Score 26; DB 22; Length 382;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMSIARL 7  
 Db 192 amsvar1 198  
 :|||

RESULT 13  
 AAB79265  
 ID AAB79265 standard; Protein; 382 AA.  
 XX  
 AC AAB79265;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:46.  
 XX  
 KW Corynebacterium glutamicum; carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100844-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB00943.  
 XX  
 XX 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031412.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031431.  
 PR 08-JUL-1999; 99DE-1031433.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 09-JUL-1999; 99DE-1032180.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 09-JUL-1999; 99US-0143208.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 31-AUG-1999; 99US-0151572.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042123.  
 PR 03-SEP-1999; 99DE-1042125.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 PI WPI; 2001-061975/07.  
 DR N-PSDB; AAF71382.  
 XX  
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes -  
 XX  
 PS Claim 20; Page 207-208; 1246pp; English.  
 XX  
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins  
 CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).  
 XX  
 SQ Sequence 382 AA;

Query Match 86.7%; Score 26; DB 22; Length 382;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMSIARL 7  
 Db 192 amsvar1 198  
 :|||

RESULT 14  
 AAW18032  
 ID AAW18032 standard; Protein; 422 AA.  
 XX  
 AC AAW18032;  
 XX  
 DT 03-SEP-1997 (first entry)  
 XX

```

DE Zebrafish retinoid X receptor RXR delta.
XX
KW Retinoid X receptor; RXR delta; zebrafish.
XX
OS Brachydanio rerio.
XX
FH Key Location/Qualifiers
FT Domain 1..89 /label= A/B
FT /note= "N-terminal domain"
FT Domain 90..156 /label= C
FT /note= "DNA-binding domain"
FT Region 157..177 /label= D
FT /note= "hinge between DNA-binding and ligand-
FT binding domains"
FT Domain 178..422 /label= E
FT /note= "ligand-binding domain"
FT Region 296..309 /note= "14-amino acid insert"
XX
CA2177642-A.
PN
XX
XX 06-DEC-1996.
XX
PF 29-MAY-1996; 96CA-2177642.
XX
PR 05-JUN-1995; 95US-0462182.
XX
XX (TOOH ) UNIV QUEENS KINGSTON.
XX
XX Jones B, Ohno C, Petkovich M;
PI
XX WPI; 1997-298660/28.
DR
DR N-PSDB; AAT67196.
XX
XX Isolated retinoid X receptor protein - having amino acid insert in
PT ligand binding domain, useful to identify specific target genes
PT implicated in retinoid responses important in disease states
XX
XX Claim 3; Page 23-25; 42pp; English.
PS
CC Novel retinoid X receptors RXR delta (AAW18032) and RXR epsilon
CC (AAW18033) exhibit a high degree of amino acid conservation with
CC other vertebrate RXRs but represent unique subtypes defined by an
CC additional 14-amino acid segment in their ligand binding domains.
CC These RXRs do not bind 9-cis retinoic acid (RA) or all-trans RA
CC with high affinity and are not activated by 9-cis RA. They are
CC able to form dimers in a manner equivalent to other RXRs. Their
CC amino acid sequences were deduced from cDNA clones (AAT67196-97)
CC obtd. from adult and post-somitogenesis zebrafish cDNA libraries.
CC The RXRs can be used to identify ligands, isolate mammalian
CC counterparts, modify retinoid activity and identify target genes
CC implicated in retinoid responses important in disease states.
XX
SQ Sequence 422 AA;

```

Query Match 86.7%; Score 26; DB 18; Length 422;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 Db 53 smsvsr1 59  
 |||:|

RESULT 15  
 AAB25565  
 ID AAB25565 standard; Protein; 697 AA.  
 XX

```

AC AAB25565;
XX
DT 27-NOV-2000 (first entry)
XX
DE Eucalyptus grandis cell signalling involved protein SEQ ID NO:940.
XX
KW Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
KW environmental change; development; cell proliferation; differentiation;
KW elongation; survival; disease resistance; nutrient metabolism.
XX
OS Eucalyptus grandis.
XX
PN WO200042171-A1.
XX
PD 20-JUL-2000.
XX
XX 11-JAN-2000; 2000WO-US00724.
XX
PR 12-JAN-1999; 99US-0228986.
PR 01-NOV-1999; 99US-0162866.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA
XX Strabala TJ, Nieuwenhuizen NJ;
PI
XX WPI; 2000-476052/41.
DR
XX Isolated polynucleotide encoding a polypeptide involved in cell
PT signalling used for generating transgenic plants with modified responses
PT to external signals -
XX
XX Claim 3; Page 518-519; 527pp; English.
PS
CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
CC pine (Pinus radiata also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC forest tree species giving long stretches of valuable knot-free clear
CC wood which can be used in solid timber furniture and veneers.
XX
SQ Sequence 697 AA;

```

Query Match 86.7%; Score 26; DB 21; Length 697;  
 Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
 Db 92 slsvarl 98  
 |:|:|

Search completed: May 8, 2002, 07:18:35  
 Job time: 48 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:17:52 ; Search time 20.74 Seconds  
(without alignments)  
15.190 Million cell updates/sec

Title: US-09-765-086-200  
Perfect score: 62  
Sequence: 1 KLAKLAKLAKLAK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCrUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	14	1 US-08-944-133-6	Sequence 6, Appl
2	62	100.0	21	1 US-08-944-133-3	Sequence 3, Appl
3	62	100.0	21	1 US-08-944-133-7	Sequence 7, Appl
4	62	100.0	21	1 US-08-944-133-26	Sequence 26, Appl
5	62	100.0	28	1 US-08-944-133-4	Sequence 4, Appl
6	62	100.0	28	1 US-08-944-133-8	Sequence 8, Appl
7	62	100.0	28	1 US-08-944-133-27	Sequence 27, Appl
8	57	91.9	14	1 US-08-944-133-25	Sequence 25, Appl
9	57	91.9	23	2 US-08-723-306-27	Sequence 27, Appl
10	57	91.9	23	2 US-08-723-306-28	Sequence 28, Appl
11	57	91.9	23	5 PCT-US96-10041-27	Sequence 27, Appl
12	57	91.9	23	5 PCT-US96-10041-28	Sequence 28, Appl
13	54	87.1	21	1 US-07-908-455A-62	Sequence 62, Appl
14	54	87.1	21	1 US-08-434-120-88	Sequence 88, Appl
15	54	87.1	21	1 US-08-465-325-86	Sequence 86, Appl
16	52	83.9	23	1 US-08-231-730A-40	Sequence 40, Appl
17	52	83.9	23	1 US-08-457-171-40	Sequence 40, Appl
18	52	83.9	23	2 US-08-505-486-45	Sequence 45, Appl
19	52	83.9	23	3 US-08-689-489C-40	Sequence 40, Appl
20	52	83.9	23	3 US-08-801-028-45	Sequence 45, Appl
21	52	83.9	23	3 US-09-340-154-45	Sequence 45, Appl
22	52	83.9	23	4 US-09-232-802A-40	Sequence 40, Appl
23	52	83.9	23	5 PCT-US95-04335-40	Sequence 40, Appl
24	52	83.9	23	5 PCT-US95-04718-40	Sequence 40, Appl
25	52	83.9	23	5 PCT-US95-09338-45	Sequence 45, Appl
26	52	83.9	23	5 PCT-US95-09339-45	Sequence 45, Appl
27	52	83.9	31	2 US-08-723-306-26	Sequence 26, Appl

Sequence 48, Appl  
Sequence 26, Appl  
Sequence 23, Appl  
Sequence 23, Appl  
Sequence 4, Appl  
Sequence 64, Appl  
Sequence 30, Appl  
Sequence 90, Appl  
Sequence 30, Appl  
Sequence 88, Appl  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 24, Appl  
Sequence 24, Appl

28 83.9 31 5 PCT-US94-12550-48  
29 52 83.9 31 5 PCT-US96-10041-26  
30 52 83.9 38 2 US-08-723-306-23  
31 52 83.9 38 5 PCT-US96-10041-23  
32 50 80.6 21 1 US-07-908-455A-4  
33 50 80.6 21 1 US-07-908-455A-64  
34 50 80.6 21 1 US-08-434-120-30  
35 50 80.6 21 1 US-08-434-120-90  
36 50 80.6 21 1 US-08-465-325-30  
37 50 80.6 21 1 US-08-465-325-88  
38 50 80.6 23 1 US-08-231-730A-24  
39 50 80.6 23 1 US-08-427-001C-24  
40 50 80.6 23 1 US-08-457-798-24  
41 50 80.6 23 1 US-08-457-171-24  
42 50 80.6 23 2 US-08-505-486-24  
43 50 80.6 23 3 US-08-689-489C-24  
44 50 80.6 23 3 US-08-801-028-24  
45 50 80.6 23 5 PCT-US94-06176-24

#### ALIGNMENTS

RESULT 1  
US-08-944-133-6  
Sequence 6, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L  
TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Rannels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Rannels, John H  
REGISTRATION NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 504 387-3221  
TELEFAX: 504 346-8049  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-6

Query Match 100.0%; Score 62; DB 1; Length 14;

Best Local Similarity 100.0%; Pred. No. 0.0029;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | | |  
Db 1 KLAKLAKKLAKLAK 14

RESULT 2  
US-08-944-133-3  
; Sequence 3, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Rannels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rannels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-3

Query Match 100.0%; Score 62; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | | |  
Db 5 KLAKLAKKLAKLAK 18

RESULT 3  
US-08-944-133-7  
; Sequence 7, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L

; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Rannels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rannels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-7

Query Match 100.0%; Score 62; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | | |  
Db 1 KLAKLAKKLAKLAK 14

RESULT 4  
US-08-944-133-26  
; Sequence 26, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Rannels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Rannels, John H  
REGISTRATION NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 504 387-3221  
TELEFAX: 504 346-8049  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-26

Query Match 100.0%; Score 62; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | | | | | | | | |  
Db 2 KLAKLAKKLAKLAK 15

RESULT 5  
US-08-944-133-4  
Sequence 4, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L  
TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Rannels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Rannels, John H  
REGISTRATION NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 504 387-3221  
TELEFAX: 504 346-8049  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-4

Query Match 100.0%; Score 62; DB 1; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.0054;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | | | | | | | | |  
Db 5 KLAKLAKKLAKLAK 18

RESULT 6  
US-08-944-133-8  
Sequence 8, Application US/08944133  
Patent No. 5789542  
GENERAL INFORMATION:  
APPLICANT: McLaughlin, Mark L  
APPLICANT: Becker, Calvin L  
TITLE OF INVENTION: Amphipathic Peptides  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John H. Rannels  
STREET: P. O. Box 2471  
CITY: Baton Rouge  
STATE: LA  
COUNTRY: USA  
ZIP: 70821-2471  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/944,133  
FILING DATE: 06-OCT-1997  
CLASSIFICATION: 5530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/789,077  
FILING DATE: 03-FEB-1997  
APPLICATION NUMBER: US/08/681,075  
FILING DATE:  
APPLICATION NUMBER: US/08/232,525  
FILING DATE: 22-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Rannels, John H  
REGISTRATION NUMBER: 33451  
REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 504 387-3221  
TELEFAX: 504 346-8049  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-944-133-8

Query Match 100.0%; Score 62; DB 1; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.0054;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | |  
Db 1 KLAKLAKKLAKLAK 14

RESULT 7  
US-08-944-133-27  
; Sequence 27, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Runnels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Runnels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-27

Query Match 100.0%; Score 62; DB 1; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.0054;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | |  
Db 2 KLAKLAKKLAKLAK 15

RESULT 8  
US-08-944-133-25  
; Sequence 25, Application US/08944133  
; Patent No. 5789542  
; GENERAL INFORMATION:  
; APPLICANT: McLaughlin, Mark L  
; APPLICANT: Becker, Calvin L  
; TITLE OF INVENTION: Amphipathic Peptides  
; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John H. Runnels  
; STREET: P. O. Box 2471  
; CITY: Baton Rouge  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70821-2471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,133  
; FILING DATE: 06-OCT-1997  
; CLASSIFICATION: 5530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/789,077  
; FILING DATE: 03-FEB-1997  
; APPLICATION NUMBER: US/08/681,075  
; FILING DATE:  
; APPLICATION NUMBER: US/08/232,525  
; FILING DATE: 22-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Runnels, John H  
; REGISTRATION NUMBER: 33451  
; REFERENCE/DOCKET NUMBER: Atty File No. 5789542 9301  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 504 387-3221  
; TELEFAX: 504 346-8049  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-944-133-25

Query Match 91.9%; Score 57; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 0.014; Mismatches 0; Indels 0; Gaps 0;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLA 13  
| | | | | | | | | |  
Db 2 KLAKLAKKLAKLA 14

RESULT 9  
US-08-723-306-27  
; Sequence 27, Application US/08723306  
; Patent No. 5856178  
; GENERAL INFORMATION:  
; APPLICANT: White PhD, Kenneth  
; APPLICANT: Morrey PhD, John  
; APPLICANT: Reed, William  
; TITLE OF INVENTION: Cassette for Expression of Lytic  
; TITLE OF INVENTION: Peptides in Mammalian Transgenic Organisms  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

```

/ FILING DATE:
/ CLASSIFICATION: 800
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Swelgert PhD, Susan E
/ REGISTRATION NUMBER: 36,289
/ REFERENCE/DOCKET NUMBER: 2549
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 8015321922
/ TELEFAX: 8015319168
/ INFORMATION FOR SEQ ID NO: 28:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 23 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: YES
/ US-08-723-306-28

```

```

RESULT 12
PCIT-US96-10041-28
; Sequence 28, Application PC/TUS9610041
; GENERAL INFORMATION:
; APPLICANT: White PhD, Kenneth
; APPLICANT: Morrey PhD, John
; APPLICANT: Reed, William
; TITLE OF INVENTION: Cassette for Expression of Lytic
; TITLE OF INVENTION: Peptides in Mammalian Transgenic
; Organisms

```



; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 88:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-434-120-88

Query Match 87.1%; Score 54; DB 1; Length 21;  
Best Local Similarity 71.4%; Pred. No. 0.053;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
|:|:|:|:|:|:|:|  
Db 5 KIAKIARKIARIAK 18

## RESULT 15

US-08-465-325-86  
; Sequence 86, Application US/08465325  
; Patent No. 5686563  
; GENERAL INFORMATION:  
; APPLICANT: Magainin Pharmaceuticals Inc.  
; APPLICANT: 5110 Campus Drive  
; APPLICANT: Plymouth Meeting, PA 19462  
; TITLE OF INVENTION: Biologically Active Peptides Having  
; TITLE OF INVENTION: N-Terminal Substitutions  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESSEE: Dunner  
; STREET: 1300 I. Street, N.W. Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,325  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/184,462  
; FILING DATE: 18-JAN-94  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE: 01-JUN-92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fordis, Jean B  
; REGISTRATION NUMBER: 32,984  
; REFERENCE/DOCKET NUMBER: 05387.0021-03000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-465-325-86

Query Match 87.1%; Score 54; DB 1; Length 21;

Best Local Similarity 71.4%; Pred. No. 0.053;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KLAKLAKKLAKLAK 14  
|:|:|:|:|:|:|:|  
Db 5 KIAKIARKIARIAK 18

Search completed: May 8, 2002, 07:19:01  
Job time: 69 sec

11



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:17:47 ; Search time 38.66 seconds  
(without alignments)  
26.824 Million cell updates/sec

Title: US-09-765-086-200  
Perfect score: 62  
Sequence: 1 KLAKLAKKLAKLAK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	14	19 AAW62925	Minimalist lytic p
2	62	100.0	14	21 AAB21900	Antimicrobial pro-
3	62	100.0	14	22 AAE06478	Synthetic anti-mic
4	62	100.0	21	11 AAR07747	Lytic peptide with
5	62	100.0	21	19 AAW62942	Minimalist lytic p
6	62	100.0	21	19 AAW62926	Minimalist lytic p
7	62	100.0	21	19 AAW62926	Minimalist lytic p
8	62	100.0	21	21 AAB21936	Homing antimicrobi
9	62	100.0	21	21 AAB21939	Homing antimicrobi
10	62	100.0	21	22 AAE06512	Prostate-homing pr
11	62	100.0	21	22 AAE06514	Homing pro-apoptot

12	62	100.0	21	22 AAE06515	Homing pro-apoptot
13	62	100.0	23	11 AAR07741	Lytic peptide with
14	62	100.0	23	11 AAB21938	Prostate homing an
15	62	100.0	23	22 AAE06513	Chimeric prostate-
16	62	100.0	23	21 AAB21940	Homing antimicrobi
17	62	100.0	25	22 AAE06517	Homing pro-apoptot
18	62	100.0	26	21 AAB21937	Homing antimicrobi
19	62	100.0	26	22 AAE06516	Homing pro-apoptot
20	62	100.0	28	19 AAW62943	Minimalist lytic p
21	62	100.0	28	19 AAW62923	Minimalist lytic p
22	62	100.0	28	19 AAW62927	Minimalist lytic p
23	57	91.9	14	19 AAW62941	Minimalist lytic p
24	57	91.9	38	11 AAR07743	Lytic peptide with
25	54	87.1	21	12 AAR13821	Amphiphilic peptid
26	54	87.1	21	14 AAR45107	Basic (positively
27	54	87.1	21	14 AAR36366	Amphiphilic ion ch
28	54	87.1	21	14 AAR31139	C-terminal subst.
29	54	87.1	21	14 AAR35357	Amphiphilic peptid
30	54	87.1	21	14 AAR39069	Biologically activ
31	54	87.1	21	15 AAR55963	Ion channel formin
32	54	87.1	21	15 AAR50423	Amphiphilic peptid
33	54	87.1	21	15 AAR50539	Amphiphilic peptid
34	54	87.1	21	15 AAR56924	Peptide which neut
35	54	87.1	21	15 AAR59041	Cancer treating, a
36	54	87.1	21	16 AAR90129	Ion-channel formin
37	54	87.1	21	16 AAR83902	Ion channel formin
38	54	87.1	21	20 AAV10725	Peptide used to ma
39	52	83.9	23	16 AAR84167	Peptide enhancer o
40	52	83.9	23	16 AAR77081	Synthetic anti-neo
41	52	83.9	23	17 AAR92416	Lytic peptide used
42	52	83.9	23	17 AAR89973	Synthetic lytic pe
43	52	83.9	38	11 AAR07737	Lytic peptide with
44	50	80.6	21	14 AAR45051	Basic (positively
45	50	80.6	21	14 AAR45108	Basic (positively

ALIGNMENTS

RESULT 1  
AAW62925  
ID AAW62925 standard; peptide; 14 AA.

AC AAW62925;

DT 02-OCT-1998 (first entry)

XX Minimalist lytic peptide.

DE Lytic peptide; channel forming peptide; antibacterial; amphipathic.

XX Synthetic.

XX US5789542-A.

XX PD 04-AUG-1998

XX PF 06-OCT-1997; 97US-0944133.

XX PR 22-APR-1994; 94US-0232525.

XX PR 22-JUL-1996; 96US-0681075.

XX PR 03-FEB-1997; 97US-0789077.

XX PR 06-OCT-1997; 97US-0944133.

XX PA (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.

XX PI Becker CL, McLaughlin ML;

XX DR WPI; 1998-446183/38.

XX PT Selective lysis of bacteria amongst mammalian cells - using 14-mer

XX PT or 21-mer lytic peptides

PS Claim 3; Column 35; 25pp; English.

XX AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal  
 CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively  
 CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha-helix, whereby the multimer is amphipathic.

XX Sequence 14 AA;

Query Match 100.0%; Score 62; DB 19; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.0029;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAKLAKKLAKLAK 14  
 Db 1 klaklakklaklak 14

RESULT 2

AA21900  
 ID AAB21900 standard; Peptide; 14 AA.

XX AAB21900;

XX 22-MAR-2001 (first entry)

DE Antimicrobial pro-apoptotic peptide #1.

XX Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
 KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;  
 KW alpha-helix.

XX Unidentified.

XX Key Location/Qualifiers

FH Misc-difference 1..14

FT /note= "Preferably D-form residues"

XX WO200042973-A2.

XX 27-JUL-2000.

XX 21-JAN-2000; 2000WO-US01602.

XX 22-JAN-1999; 99US-0235902.

XX (BURN-) BURNHAM INST.

XX Ellerby HM, Bredeesen DE, Pasqualini R, Ruoslahti EI;

XX WPI; 2000-499174/44.

XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
 PT selectively homes to a mammalian cell type or tissue linked to an  
 PT antimicrobial peptide, useful for the treatment of prostate cancer -

XX Claim 4; Page 104; 118pp; English.

XX The present invention relates to homing pro-apoptotic conjugates,  
 CC comprising of a tumour homing molecule that selectively homes to a  
 CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
 CC homing pro-apoptotic conjugates are selectively internalised by the  
 CC mammalian cell type or tissue and exhibits high toxicity, especially to  
 CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
 CC toxicity when not linked to the tumor homing molecule. The conjugates are

CC useful for the treatment of cancer e.g. Kaposi's sarcoma, breast and  
 CC prostate cancer or melanoma. The present sequence is one such  
 CC antimicrobial peptide, which can be conjugated to a homing peptide to  
 CC make the homing pro-apoptotic conjugates of the present invention. The  
 CC present sequence has an amphipathic alpha-helical structure.

XX Sequence 14 AA;

Query Match 100.0%; Score 62; DB 21; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.0029;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAKLAKKLAKLAK 14  
 Db 1 klaklakklaklak 14

RESULT 3

AAE06478  
 ID AAE06478 standard; peptide; 14 AA.

XX AAE06478;

XX 25-SEP-2001 (first entry)

XX Synthetic anti-microbial peptide #1.

XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
 KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
 KW cytostatic.

XX Synthetic.

XX WO200153342-A1.

XX 26-JUL-2001.

XX 16-JAN-2001; 2001WO-US01362.

XX 21-JAN-2000; 2000US-0489582.

XX (BURN-) BURNHAM INST.

XX Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;

XX WPI; 2001-451901/48.

XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
 PT prostate cancer, comprises a prostate-homing peptide linked to an  
 PT antimicrobial peptide -

XX Claim 4; Page 103; 176pp; English.

XX The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used  
 CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is a synthetic  
 CC anti-microbial peptide having an amphipathic-alpha helical structure.  
 CC This peptide is linked to a prostate-homing peptide to generate a  
 CC chimeric prostate-homing pro-apoptotic peptide.

XX Sequence 14 AA;

Query Match 100.0%; Score 62; DB 22; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.0029;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0;

QY 1 KLAKLAKKLAKLAK 14  
 DB 1 klaklakklaklak 14

RESULT  
 AAR07747

ID AAR07747 standard; protein; 21 AA.

XX AAR07747;

AC AAR07747;

XX 22-FEB-1991 (first entry)

DE Lytic peptide with proliferative activity.

XX Lysis; neoplastic cells; microbial infections; HIV; P.falciparum;

KW wound healing; adjuvant; hBFSP.

XX Synthetic

PN W09012866-A

XX 01-NOV-1990.

ED 10-APR-1990; 90WO-US01945.

XX 10-APR-1989; 89US-0336181.

XX (LOU ) LOUISIANA STATE UNIV.

XX Jaynes JM;

XX WPI; 1990-348469/46.

XX New lytic polypeptide(s) with proliferative activity - are

PT alpha-helical peptide(s) having aligned amphipathy for treating

PT microbial infections and lysing cancer cells

PS Claim 24; page 42; 57pp; English.

CC This peptide is an analogue of a known lytic peptide. It comprises  
 CC an alpha helical conformation of amino acids. It is effective at  
 CC lysing e.g. gram-positive and -negative bacteria and mammalian neo-  
 CC plastic cells, cells infected with intracellular pathogenic micro-  
 CC organisms such as HIV. It stimulates the proliferation of fibro-  
 CC blasts and lymphocytes and can be used in wound-healing.  
 CC See also: AAR07734-41, AAR07743-46 and AAR07748-51.

XX Sequence 21 AA;

Query Match 100.0%; Score 62; DB 11; Length 21;

Best Local Similarity 100.0%; Pred. No. 0.0042;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14

DB 3 klaklakklaklak 16

RESULT 5

ID AAW62942 standard; peptide; 21 AA.

XX AAW62942;

XX 02-OCT-1998 (first entry)

DE Minimalist lytic peptide.

XX Lytic peptide; channel forming peptide; antibacterial; amphipathic.

KW Lytic peptide; channel forming peptide; antibacterial; amphipathic.

XX

OS Synthetic.

XX US5789542-A.

XX 04-AUG-1998.

XX 06-OCT-1997; 97US-0944133.

XX 22-APR-1994; 94US-0232525.

XX 22-JUL-1996; 96US-0681075.

XX 03-FEB-1997; 97US-0789077.

XX 06-OCT-1997; 97US-0944133.

XX (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.

XX Becker CL, McLaughlin ML;

XX WPI; 1998-446183/38.

XX Selective lysis of bacteria amongst mammalian cells - using 14-mer  
 or 21-mer lytic peptides

XX Disclosure; Column 6; 25pp; English.

XX AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal  
 CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively  
 CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha-helix, whereby the multimer is amphipathic.

XX Sequence 21 AA;

Query Match 100.0%; Score 62; DB 19; Length 21;

Best Local Similarity 100.0%; Pred. No. 0.0042;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14

DB 2 klaklakklaklak 15

RESULT 6

ID AAW62922 standard; peptide; 21 AA.

XX AAW62922;

XX 02-OCT-1998 (first entry)

XX Minimalist lytic peptide.

XX Lytic peptide; channel forming peptide; antibacterial; amphipathic.

XX Synthetic.

XX US5789542-A.

XX 04-AUG-1998.

XX 06-OCT-1997; 97US-0944133.

XX 22-APR-1994; 94US-0232525.

XX 22-JUL-1996; 96US-0681075.

XX 03-FEB-1997; 97US-0789077.

XX 06-OCT-1997; 97US-0944133.

PA (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.  
 XX Becker CL, McLaughlin ML;  
 PI WPI; 1998-446183/38.  
 DR Selective lysis of bacteria amongst mammalian cells - using 14-mer  
 XX or 21-mer lytic peptides  
 PT Disclosure; Column 5; 25pp; English.  
 PS  
 XX  
 CC AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal  
 CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively  
 CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha -helix, whereby the multimer is amphipathic.  
 XX  
 SQ Sequence 21 AA;

Query Match 100.0%; Score 62; DB 19; Length 21;  
 Best Local Similarity 100.0%; Pred. NO. 0.0042;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAIAKLAIAKLAIAK 14  
 DB 5 KLAIAKLAIAKLAIAK 18  
 |||||  
 |||||

RESULT 7  
 AAW62926  
 ID AAW62926 standard; peptide; 21 AA.  
 XX  
 AC AAW62926;  
 XX  
 DT 02-OCT-1998 (first entry)  
 XX  
 DE Minimalist lytic peptide.  
 XX  
 KW Lytic peptide; channel forming peptide; antibacterial; amphipathic.  
 XX  
 OS Synthetic.  
 XX  
 PN US5789542-A.  
 XX  
 PD 04-AUG-1998.  
 XX  
 PF 06-OCT-1997; 97US-0944133.  
 XX  
 PR 22-APR-1994; 94US-0232525.  
 PR 22-JUL-1996; 96US-0681075.  
 PR 03-FEB-1997; 97US-0789077.  
 PR 06-OCT-1997; 97US-0944133.  
 XX  
 PA (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.  
 PI Becker CL, McLaughlin ML;  
 XX WPI; 1998-446183/38.  
 DR Selective lysis of bacteria amongst mammalian cells - using 14-mer  
 XX or 21-mer lytic peptides  
 PT Disclosure; Column 5; 25pp; English.  
 PS  
 XX  
 CC AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal

CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively  
 CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha -helix, whereby the multimer is amphipathic.  
 XX  
 SQ Sequence 21 AA;

Query Match 100.0%; Score 62; DB 19; Length 21;  
 Best Local Similarity 100.0%; Pred. NO. 0.0042;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAIAKLAIAKLAIAK 14  
 DB 1 KLAIAKLAIAKLAIAK 14  
 |||||  
 |||||

RESULT 8  
 AAB21936  
 ID AAB21936 standard; Peptide; 21 AA.  
 XX  
 AC AAB21936;  
 XX  
 DT 22-MAR-2001 (first entry)  
 XX  
 DE Homing antimicrobial pro-apoptotic conjugate #1.  
 XX  
 KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
 KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;  
 KW alpha-helix; human.  
 XX  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 8..21 /note= "Preferably D-form residues"  
 FT  
 XX WO200042973-A2.  
 XX  
 XX 27-JUL-2000.  
 XX  
 XX 21-JAN-2000; 2000WO-US01602.  
 XX  
 XX 22-JAN-1999; 99US-0235902.  
 XX  
 XX (BURN-) BURNHAM INST.  
 XX  
 XX Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;  
 XX WPI; 2000-499174/44.  
 DR  
 XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
 XX selectively homes to a mammalian cell type or tissue linked to an  
 XX antimicrobial peptide, useful for the treatment of prostate cancer -  
 XX Claim 13; Page 105; 118pp; English.  
 XX  
 CC The present invention relates to homing pro-apoptotic conjugates,  
 CC comprising of a tumour homing molecule that selectively homes to a  
 CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
 CC homing pro-apoptotic conjugates are selectively internalised by the  
 CC mammalian cell type or tissue and exhibits high toxicity, especially to  
 CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
 CC toxicity when not linked to the tumor homing molecule. In addition, the  
 CC antimicrobial peptide has an amphipathic alpha-helical structure. The  
 CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,  
 CC breast and prostate cancer or melanoma. The present sequence is one such

CC homing pro-apoptotic conjugate.

XX Sequence 21 AA;

Query Match 100.0%; Score 62; DB 21; Length 21;

Best Local Similarity 100.0%; Pred. No. 0.0042;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | |  
DB 8 klaklakklaklak 21

RESULT 9

AAB21939

ID AAB21939 standard; Peptide; 21 AA.

XX

AC AAB21939;

XX

DT 22-MAR-2001 (first entry)

XX

DE Homing antimicrobial pro-apoptotic conjugate #3.

XX

KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;

KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;

KW alpha-helix; human.

XX

OS Chimeric - Homo sapiens.

OS Chimeric - Unidentified.

XX

FH Key Location/Qualifiers

FT Misc-difference 8..21 /note= "Preferably D-form residues"

FT

XX

PN WO200042973-A2.

XX

PD 27-JUL-2000.

XX

PF 21-JAN-2000; 2000WO-US01602.

XX

PR 22-JAN-1999; 99US-0235902.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;

XX

DR WPI; 2000-499174/44.

XX

PT Homing pro-apoptotic conjugate comprising a tumor homing molecule that

selectively homes to a mammalian cell type or tissue linked to an

antimicrobial peptide, useful for the treatment of prostate cancer -

XX

PS Disclosure; Page 8; 118pp; English.

XX

CC The present invention relates to homing pro-apoptotic conjugates,

comprising of a tumour homing molecule that selectively homes to a

mammalian cell type or tissue, linked to an antimicrobial peptide. The

homing pro-apoptotic conjugates are selectively internalised by the

mammalian cell type or tissue and exhibits high toxicity, especially to

angiogenic vasculature. The antimicrobial peptide has low mammalian cell

toxicity when not linked to the tumor homing molecule. In addition, the

antimicrobial peptide has an amphipathic alpha-helical structure. The

conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,

breast and prostate cancer or melanoma. The present sequence is one such

homing pro-apoptotic conjugate.

XX

SQ Sequence 21 AA;

Query Match

Best Local Similarity 100.0%; Score 62; DB 21; Length 21;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
| | | | | | | | | | | | | | |  
DB 8 klaklakklaklak 21

RESULT 10

AAE06512

ID AAE06512 standard; peptide; 21 AA.

XX

AC AAE06512;

XX

DT 25-SEP-2001 (first entry)

XX

DE Prostate-homing pro-apoptotic peptide, HPP-1.

XX

KW Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;

antimicrobial peptide; prostate cancer; breast tumour homing molecule;

cytostatic; HPP-1.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Domain 1..5

FT Disulfide-bond 1..5

FT Domain 6..7

FT /label= Coupling\_domain

FT /note= "Glycinyglycine bridge"

FT Domain 8..21

FT /label= Membrane\_disrupting\_domain

FT /note= "Antimicrobial peptide; This region forms an

amphipathic helix and is useful in imparting increased

stability of the conjugate in vivo"

FT

XX

PN WO200153342-A1.

XX

PD 26-JUL-2001.

XX

PF 16-JAN-2001; 2001WO-US01362.

XX

PR 21-JAN-2000; 2000US-0489582.

XX

PA (BURN-) BURNHAM INST.

XX

PI Ruoslahti EI, Pasqualini R, Arap W, Bredesen DE, Ellerby HM;

XX

DR WPI; 2001-451901/48.

XX

PT Novel chimeric prostate-homing pro-apoptotic peptide, used to treat

prostate cancer, comprises a prostate-homing peptide linked to an

antimicrobial peptide -

XX

PS Example 2; Fig 1; 176pp; English.

XX

CC The patent discloses novel chimeric prostate-homing pro-apoptotic

peptide which comprises a prostate-homing peptide linked to an

antimicrobial peptide, where the chimeric peptide is selectively

internalised by and exhibits high toxicity to prostate tissue and

where the antimicrobial peptide has low mammalian cell toxicity when

not linked to prostate-homing peptide. The chimeric peptide is used

to direct an antimicrobial peptide in vivo to a prostate cancer, to

induce selective toxicity in vivo in a prostate cancer, and to treat

a patient with prostate cancer. The present sequence is chimeric

prostate-homing pro-apoptotic peptide, HPP-1.

XX

SQ Sequence 21 AA;

Query Match

Best Local Similarity 100.0%; Score 62; DB 22; Length 21;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAKLAKKLAKLAK 14  
 Db 8 klaklakklaklak 21

RESULT 11  
 AAE06514  
 ID AAE06514 standard; peptide; 21 AA.

AC AAE06514;  
 XX  
 XX 25-SEP-2001 (first entry)

DE Homing pro-apoptotic peptide #1.

KW Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
 KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
 KW cytostatic.

XX Unidentified.

XX Key Location/Qualifiers  
 FH Domain 1..5  
 FT /label= Homing\_domain  
 FT Domain 6..7  
 FT /label= Coupling\_domain  
 FT /note= "Glycylglycine bridge"  
 FT Domain 8..21  
 FT /label= Antimicrobial\_peptide

XX WO200153342-A1.

PN 26-JUL-2001.

XX 16-JAN-2001; 2001WO-US01362.

XX 21-JAN-2000; 2000US-0489582.

XX (BURN-) BURNHAM INST.

XX Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;

XX WPI; 2001-451901/48.

XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
 PT prostate cancer, comprises a prostate-homing peptide linked to an  
 PT antimicrobial peptide -

XX Example 2; Page 80; 176pp; English.

XX The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used  
 CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is a homing pro-  
 CC apoptotic peptide.

XX Sequence 21 AA;

Query Match 100.0%; Score 62; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.0042;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAKLAKKLAKLAK 14  
 Db 8 klaklakklaklak 21

RESULT 12  
 AAE06515

ID AAE06515 standard; peptide; 21 AA.

XX AAE06515;

XX 25-SEP-2001 (first entry)

DE Homing pro-apoptotic peptide #2.

XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
 KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
 KW cytostatic.

XX Unidentified.

XX Key Location/Qualifiers  
 FH Domain 1..5  
 FT /label= Homing\_domain  
 FT Domain 6..7  
 FT /label= Coupling\_domain  
 FT /note= "Glycylglycine bridge"  
 FT Domain 8..21  
 FT /label= Antimicrobial\_peptide

XX WO200153342-A1.

PN 26-JUL-2001.

XX 16-JAN-2001; 2001WO-US01362.

XX 21-JAN-2000; 2000US-0489582.

XX (BURN-) BURNHAM INST.

XX Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;

XX WPI; 2001-451901/48.

XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
 PT prostate cancer, comprises a prostate-homing peptide linked to an  
 PT antimicrobial peptide -

XX Example 2; Page 80; 176pp; English.

XX The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used  
 CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is a homing pro-  
 CC apoptotic peptide.

XX Sequence 21 AA;

Query Match 100.0%; Score 62; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.0042;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLAKLAKKLAKLAK 14  
 Db 8 klaklakklaklak 21

RESULT 13  
 AAR07741

ID AAR07741 standard; protein; 23 AA.

XX AAR07741;

```

XX 22-FEB-1991 (first entry)
DT Lytic peptide with proliferative activity.
DE
XX Lysis; neoplastic cells; microbial infections; HIV; P.falci-parum;
KW wound healing; adjuvant; hBFSF.
KW
XX synthetic.
OS
XX WO9012866-A.
PN
XX
XX 01-NOV-1990.
PD
XX
XX 10-APR-1990; 90WO-US01945.
PF
XX
XX 10-APR-1989; 89US-0336181.
PR
XX (LOU ) LOUISIANA STATE UNIV.
PA
XX Jaynes JM;
XX
XX WPI; 1990-348469/46.
DR
XX New lytic polypeptide(s) with proliferative activity - are
PT alpha-helical peptide(s) having aligned amphipathy for treating
PT microbial infections and lysing cancer cells
XX
XX Claim 17; page 41; 57pp; English.
PS
XX This peptide is an analogue of a known lytic peptide. It comprises
CC an alpha helical conformation of amino acids. It is effective at
CC lysing e.g. gram-positive and -negative bacteria and mammalian neo-
CC plastic cells, cells infected with intracellular pathogenic micro-
CC organisms such as HIV. It stimulates the proliferation of fibro-
CC blasts and lymphocytes and can be used in wound healing.
CC See also AAR07734-40 and AAR07743-51.
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 62; DB 11; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLAKLAKKLAKLAK 14
Db 4 klaklakklaklak 17
RESULT 14
AAB21938
ID AAB21938 standard; Peptide; 23 AA.
XX
XX AAB21938;
AC
XX
XX 22-MAR-2001 (first entry)
DT
XX Prostate homing antimicrobial pro-apoptotic conjugate.
DE
XX Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;
KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;
KW alpha-helix; human.
XX
XX Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
FH Misc-difference 10..23
FT /note= "Preferably D-form residues"
XX
XX WO200042973-A2.
PN
XX

```

```

PD 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-US01602.
PF
XX
XX 22-JAN-1999; 99US-0235902.
PR
XX (BURN-) BURNHAM INST.
PA
XX
XX Ellerby HM, Bredeesen DE, Pasqualini R, Ruoslahti EI;
PI
XX WPI; 2000-499174/44.
DR
XX
XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that
XX selectively homes to a mammalian cell type or tissue linked to an
XX antimicrobial peptide, useful for the treatment of prostate cancer -
PT
XX Claim 29; Page 108; 118pp; English.
PS
XX The present invention relates to homing pro-apoptotic conjugates,
XX comprising of a tumour homing molecule that selectively homes to a
XX mammalian cell type or tissue, linked to an antimicrobial peptide. The
XX homing pro-apoptotic conjugates are selectively internalised by the
XX mammalian cell type or tissue and exhibits high toxicity, especially to
XX angiogenic vasculature. The antimicrobial peptide has low mammalian cell
XX toxicity when not linked to the tumor homing molecule. In addition, the
XX antimicrobial peptide has an amphipathic alpha-helical structure. The
XX conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,
XX breast and prostate cancer or melanoma. The present sequence is one such
XX prostate homing pro-apoptotic conjugate.
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 62; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLAKLAKKLAKLAK 14
Db 10 klaklakklaklak 23
RESULT 15
AAE06513
ID AAE06513 standard; peptide; 23 AA.
XX
XX AAE06513;
AC
XX
XX 25-SEP-2001 (first entry)
DT
XX Chimeric prostate-homing pro-apoptotic peptide.
DE
XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;
KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;
KW cytostatic.
XX
XX Unidentified.
OS
XX
XX Key Location/Qualifiers
FH Domain 1..7
FT /label= Prostate_homing_domain
FT Domain 8..9
FT /label= Coupling_domain
FT /note= "Glycinyglycine bridge"
FT Domain 10..23
FT /label= Antimicrobial_peptide
XX
XX WO200153342-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 16-JAN-2001; 2001WO-US01362.
PF
XX
XX

```

PR 21-JAN-2000; 2000US-0489582.  
 XX  
 PA (BURN-) BURNHAM INST.  
 XX  
 PI Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;  
 XX  
 DR WPI; 2001-451901/48.  
 XX  
 PT Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
 PT prostate cancer, comprises a prostate-homing peptide linked to an  
 PT antimicrobial peptide -  
 XX  
 PS Claim 6; Page 103; 176pp; English.  
 XX  
 CC The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used  
 CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is a chimeric  
 CC prostate-homing pro-apoptotic peptide.  
 XX  
 SQ Sequence 23 AA;

Query Match 100.0%; Score 62; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLAKLAKKLAKLAK 14  
 DB 10 klaklakklaklak 23

Search completed: May 8, 2002, 07:18:34  
 Job time: 47 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 07:19:01 ; Search time 20.74 seconds  
(without alignments)  
7.595 Million cell updates/sec

Title: US-09-765-086-207  
Perfect score: 30  
Sequence: 1 SMSIARL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2.6/ptodata/2/iaa/PTUS\_COMB.pep.\*
- 6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	7	4	US-09-258-754-21
2	30	100.0	7	4	US-09-042-107-21
3	23	76.7	118	3	US-09-045-765A-5
4	23	76.7	439	4	US-09-413-814-13
5	23	76.7	476	2	US-08-850-880-2
6	23	76.7	476	2	US-08-944-916-2
7	23	76.7	476	2	US-08-814-877-2
8	23	76.7	554	4	US-08-180-371-2
9	23	76.7	700	4	US-09-413-814-68
10	22	73.3	175	1	US-08-102-691-2
11	22	73.3	177	6	5430019-2
12	22	73.3	200	3	US-09-248-335-72
13	22	73.3	205	6	5175383-6
14	22	73.3	206	1	US-08-102-691-1
15	22	73.3	206	1	US-08-439-725A-13
16	22	73.3	206	1	US-08-464-590A-15
17	22	73.3	206	1	US-08-441-629-10
18	22	73.3	206	1	US-08-462-169B-12
19	22	73.3	206	2	US-08-207-412B-10
20	22	73.3	206	2	US-08-867-471-13
21	22	73.3	206	2	US-08-438-439C-9
22	22	73.3	206	2	US-08-951-822-26
23	22	73.3	206	3	US-09-103-079-12
24	22	73.3	206	3	US-08-705-245-13
25	22	73.3	206	3	US-08-718-904-13
26	22	73.3	206	3	US-09-023-082A-13
27	22	73.3	206	3	US-08-776-207-10

28	73.3	206	3	US-09-093-585-15	Sequence 15, Appl
29	73.3	206	5	PCT-US95-09172-10	Sequence 10, Appl
30	73.3	331	4	US-08-810-712-24	Sequence 24, Appl
31	73.3	335	1	US-08-348-891A-4	Sequence 4, Appl
32	73.3	335	2	US-08-905-817-4	Sequence 4, Appl
33	73.3	497	1	US-08-513-841-2	Sequence 2, Appl
34	73.3	497	2	US-08-696-834-2	Sequence 2, Appl
35	73.3	497	2	US-08-942-673-2	Sequence 2, Appl
36	73.3	497	4	US-09-118-317-2	Sequence 2, Appl
37	73.3	551	2	US-08-468-249A-20	Sequence 20, Appl
38	73.3	933	2	US-08-313-200-1	Sequence 1, Appl
39	73.3	933	5	PCT-US93-03837-1	Sequence 1, Appl
40	73.3	1091	3	US-08-986-485-5	Sequence 5, Appl
41	73.3	1423	4	US-08-810-712-10	Sequence 10, Appl
42	73.3	4544	1	US-08-469-486-52	Sequence 52, Appl
43	73.3	4544	2	US-08-469-658-52	Sequence 52, Appl
44	70.0	14	4	US-09-041-886-49	Sequence 49, Appl
45	70.0	14	4	US-09-041-886-55	Sequence 55, Appl

ALIGNMENTS

RESULT 1  
US-09-258-754-21  
; Sequence 21, Application US/09258754  
; Patent No. 6174687  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; APPLICANT: Rajotte, Daniel  
; TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using  
; FILE REFERENCE: P-LJ 3443  
; CURRENT APPLICATION NUMBER: US/09/258,754  
; CURRENT FILING DATE: 1999-02-26  
; EARLIER FILING DATE: 1998-03-13  
; NUMBER OF SEQ ID NOS: 452  
; SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-258-754-21

Query Match 100.0%; Score 30; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
Db 1 SMSIARL 7

RESULT 2  
US-09-042-107-21  
; Sequence 21, Application US/09042107  
; Patent No. 6232287  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: Molecules that Home to Various Selected Organs or  
; TITLE OF INVENTION: Tissues  
; FILE REFERENCE: P-LJ 2892  
; CURRENT APPLICATION NUMBER: US/09/042,107  
; CURRENT FILING DATE: 1998-03-13  
; NUMBER OF SEQ ID NOS: 436  
; SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21

```
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-042-107-21

Query Match          100.0%; Score 30; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 1 SMSIARL 7

RESULT 3
US-09-045-764A-5
; Sequence 5, Application US/09045764A
; Patent No. 6127178
; GENERAL INFORMATION:
; APPLICANT: Israel, Mark A.
; APPLICANT: Florio, Monica
; TITLE OF INVENTION: Apoptotic Peptides
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,764A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCSF98-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 118 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-045-764A-5

Query Match          76.7%; Score 23; DB 3; Length 118;
Best Local Similarity 83.3%; Pred. No. 78;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIAR 6
Db 24 SLSIAR 29

RESULT 4
US-09-413-814-13
; Sequence 13, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
```

```
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M.
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-13

Query Match          76.7%; Score 23; DB 4; Length 439;
Best Local Similarity 71.4%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7
Db 260 SLSIGRL 266

RESULT 5
US-08-850-880-2
; Sequence 2, Application US/08850880
; Patent No. 5925545
; GENERAL INFORMATION:
; APPLICANT: Reznikoff, William S
; APPLICANT: Gorysin, Igor Y
; APPLICANT: Zhou, Hong
; TITLE OF INVENTION: System for In Vitro Transposition
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,880
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berson, Bennett J
; REGISTRATION NUMBER: 37094
; REFERENCE/DOCKET NUMBER: 960296.94142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608/251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 amino acids
; TYPE: amino acid
```

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-850-880-2

Query Match 76.7%; Score 23; DB 2; Length 476;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSTARL 7  
|:||||  
Db 427 MAIARL 432

RESULT 6  
US-08-944-916-2  
; Sequence 2, Application US/08944916  
; Patent No. 5948622  
; GENERAL INFORMATION:  
; APPLICANT: Reznikoff, William S  
; APPLICANT: Goryshin, Igor Y  
; APPLICANT: York, Dona L  
; APPLICANT: Zhou, Hong  
; TITLE OF INVENTION: System for In Vitro Transposition  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,916  
; FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/814,877  
; FILING DATE: 09-SEP-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/850,880  
; FILING DATE: 02-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berson, Bennett J  
; REGISTRATION NUMBER: 37094  
; REFERENCE/DOCKET NUMBER: 960296.94916  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608/251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 476 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-944-916-2

Query Match 76.7%; Score 23; DB 2; Length 476;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSTARL 7  
|:||||  
Db 427 MAIARL 432

RESULT 7

US-08-814-877-2  
; Sequence 2, Application US/08814877  
; Patent No. 5965443  
; GENERAL INFORMATION:  
; APPLICANT: Goryshin, Igor Y  
; APPLICANT: Reznikoff, William S  
; TITLE OF INVENTION: System for In Vitro Transposition  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/814,877  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berson, Bennett J  
; REGISTRATION NUMBER: 37094  
; REFERENCE/DOCKET NUMBER: 960296.94142  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608/251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 476 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-814-877-2

Query Match 76.7%; Score 23; DB 2; Length 476;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MSTARL 7  
|:||||  
Db 427 MAIARL 432

RESULT 8  
US-08-180-371-2  
; Sequence 2, Application US/08180371  
; Patent No. 6254861  
; GENERAL INFORMATION:  
; APPLICANT: Choudhury, Chandra  
; TITLE OF INVENTION: Hematopoietic Growth Factor Derived  
; TITLE OF INVENTION: from T Lymphocytes and Methods of Use Therefor  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/180,371

; FILING DATE: 12-JAN-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 906866  
; FILING DATE: 01 July 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gruber, Lewis S.  
; REGISTRATION NUMBER: 30,060  
; REFERENCE/DOCKET NUMBER: 27620/31668  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEFAX: 25-3856  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 554 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-180-371-2

Query Match 76.7%; Score 23; DB 4; Length 554;  
Best Local Similarity 85.7%; Pred. No. 4.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
| | | | |  
Db 166 SHSIARL 172

RESULT 9  
US-09-413-814-68  
; Sequence 68; Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; EARLIER FILING DATE: 1999-10-07  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 68  
; LENGTH: 700  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-68

Query Match 76.7%; Score 23; DB 4; Length 700;  
Best Local Similarity 71.4%; Pred. No. 5.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
| | | | |  
Db 358 SISMARL 364

RESULT 10

US-08-102-691-2  
; Sequence 2; Application 08/102691  
; Patent No. 5559093  
; GENERAL INFORMATION:  
; APPLICANT: YOSHITOMI, Sumie  
; APPLICANT: KURORAWA, Tsutomu  
; APPLICANT: IGARASHI, Koichi  
; TITLE OF INVENTION: PLATELET-INCREASING AGENT  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: 08/102,691  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: EISENSTEIN, Ronald I  
; REGISTRATION NUMBER: 30628  
; REFERENCE/DOCKET NUMBER: 42835  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)523-3400  
; TELEFAX: (617)523-6440  
; TELEFAX: 200921 STRE DR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 175 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-102-691-2

Query Match 73.3%; Score 22; DB 1; Length 175;  
Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
| | | | |  
Db 21 ALSLARL 27

RESULT 11  
5430019-2  
; Patent No. 5430019  
; APPLICANT: ROGERS, DAVID T.; WOLFMAN, NEIL M.; SEEHRA, JASBIR S.  
; TITLE OF INVENTION: HOMOGENEOUS K-FGF AND USE OF THE SAME  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/898,051  
; FILING DATE: 12-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 438,278  
; FILING DATE: 16-NOV-1989  
; SEQ ID NO: 2  
; LENGTH: 177  
5430019-2

Query Match 73.3%; Score 22; DB 6; Length 177;  
Best Local Similarity 57.1%; Pred. No. 2.2e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSTARL 7  
: : : : :  
Db 23 ALSLARL 29

## RESULT 12

US-09-248-335-72  
; Sequence 72, Application US/09248335  
; Patent No. 6096504  
; GENERAL INFORMATION:  
; APPLICANT: MCGONIGLE, BRIAN  
; APPLICANT: O'KEEF, DANIEL  
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES  
; FILE REFERENCE: CL-1128-A  
; CURRENT APPLICATION NUMBER: US/09/248,335  
; CURRENT FILING DATE: 1999-02-10  
; EARLIER APPLICATION NUMBER: 08/924,759  
; EARLIER FILING DATE: 1997-September-05  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: Microsoft Word Version 7.0A  
; SEQ ID NO 72  
; LENGTH: 200  
; TYPE: PRT  
; ORGANISM: maize  
US-09-248-335-72

Query Match 73.3%; Score 22; DB 3; Length 200;  
Best Local Similarity 57.1%; Pred. No. 2.5e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSTARL 7  
: : : : :  
Db 184 SLTLARL 190

## RESULT 13

5175383-6  
; Patent No. 5175383  
; APPLICANT: LEDER, PHILIP;MULLER, WILLIAM J.  
; TITLE OF INVENTION: ANIMAL MODEL FOR BENIGN PROSTATIC DISEASE  
; NUMBER OF SEQUENCES: 9  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/312,641  
; FILING DATE: 17-FEB-1989  
; SEQ ID NO:6:  
; LENGTH: 205  
5175383-6

Query Match 73.3%; Score 22; DB 6; Length 205;  
Best Local Similarity 57.1%; Pred. No. 2.6e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSTARL 7  
: : : : :  
Db 52 ALSLARL 58

## RESULT 14

US-08-102-691-1  
; Sequence 1, Application 08/102691  
; Patent No. 5559093  
; GENERAL INFORMATION:  
; APPLICANT: YOSHITOMI, Sumie  
; APPLICANT: KUROKAWA, Tsutomu  
; APPLICANT: IGARASHI, Koichi  
; TITLE OF INVENTION: PLATELET-INCREASING AGENT  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 Water Street

; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: 08/102,691  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: EISENSTEIN, Ronald I  
; REGISTRATION NUMBER: 30628  
; REFERENCE/DOCKET NUMBER: 42835  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)523-3400  
; TELEFAX: (617)523-6440  
; TELEX: 200921 STRE UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 206 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-102-691-1

Query Match 73.3%; Score 22; DB 1; Length 206;  
Best Local Similarity 57.1%; Pred. No. 2.6e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSTARL 7  
: : : : :  
Db 52 ALSLARL 58

## RESULT 15

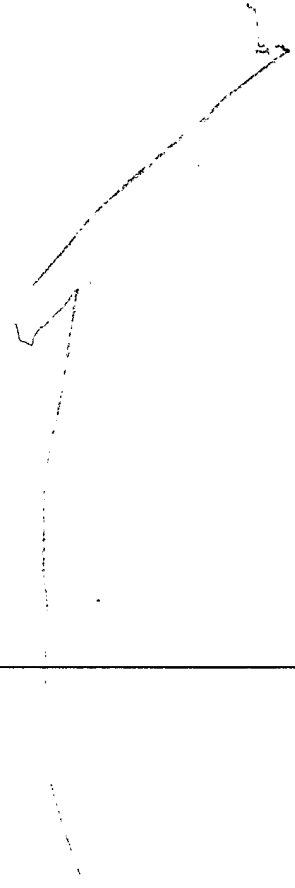
US-08-439-725A-13  
; Sequence 13, Application US/08439725A  
; Patent No. 5693775  
; GENERAL INFORMATION:  
; APPLICANT: Nathans, Jeremy  
; APPLICANT: Smallwood, Philip M.  
; APPLICANT: Macke, Jennifer P.  
; TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR HOMOLOGOUS  
; TITLE OF INVENTION: FACTOR-1 (FHF-1) AND METHODS OF USE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/439,725A  
; APPLICATION NUMBER: US/08/439,725A  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07265/047001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 617/678-5099

; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 206 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-439-725A-13

Query Match 73.3%; Score 22; DB 1; Length 206;  
Best Local Similarity 57.1%; Pred. No. 2.6e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMSIARL 7  
: : : : :  
Db 52 ALSIARL 58

Search completed: May 8, 2002, 07:19:02  
Job time: 70 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 13:31:42 ; Search time 23.87 seconds  
(without alignments)  
71.374 Million cell updates/sec

Title: CHIMERA

Perfect score: 104

Sequence: 1 SMSTARLGKGLAKLAKLAKLAK 23

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	23	21	prostate homing an
2	104	100.0	23	22	Chimeric prostate-
3	78	75.0	21	21	Homing antimicrobi
4	78	75.0	21	22	Prostate-homing pr
5	74	71.2	21	21	Homing antimicrobi
6	74	71.2	21	22	Homing pro-apoptot
7	74	71.2	21	22	Homing pro-apoptot
8	74	71.2	25	21	Homing antimicrobi
9	74	71.2	25	22	Homing pro-apoptot
10	74	71.2	26	21	Homing antimicrobi
11	74	71.2	26	22	Homing pro-apoptot

12	69.2	21	19	AAW62926	Minimalist lytic p
13	69.2	28	19	AAW62943	Minimalist lytic p
14	69.2	28	19	AAW62923	Minimalist lytic p
15	69.2	28	19	AAW62927	Minimalist lytic p
16	64.4	21	19	AAW62942	Minimalist lytic p
17	66	63.5	21	AAW62925	Minimalist lytic p
18	62	59.6	14	AAW62925	Minimalist lytic p
19	62	59.6	14	AAW62925	Antimicrobial pro-
20	62	59.6	14	AAW62925	Synthetic anti-mic
21	62	59.6	21	AAW62925	Lytic peptide with
22	62	59.6	21	AAW62925	Basic (positively
23	62	59.6	21	AAW62925	Amphiphilic ion ch
24	62	59.6	21	AAW62925	C-terminal substd.
25	62	59.6	21	AAW62925	Amphiphilic peptid
26	62	59.6	21	AAW62925	Biologically activ
27	62	59.6	21	AAW62925	Ion channel formi
28	62	59.6	21	AAW62925	Amphiphilic pepti
29	62	59.6	21	AAW62925	Amphiphilic peptid
30	62	59.6	21	AAW62925	Peptide which neut
31	62	59.6	21	AAW62925	Cancer treating, a
32	62	59.6	21	AAW62925	(KLAKGLA)3 peptide
33	62	59.6	21	AAW62925	Ion channel formi
34	62	59.6	21	AAW62925	Peptide used to ma
35	62	59.6	23	11	Lytic peptide with
36	61	58.7	38	11	Lytic peptide with
37	61	58.7	42	14	Biologically activ
38	60	57.7	21	12	Amphiphilic peptid
39	60	57.7	21	12	PGIA peptide compo
40	60	57.7	21	14	Basic (positively
41	60	57.7	21	14	Basic (positively
42	60	57.7	21	14	Amphiphilic ion ch
43	60	57.7	21	14	Amphiphilic ion ch
44	60	57.7	21	14	C-terminal substd.
45	60	57.7	21	17	PGIA-derived antim

#### ALIGNMENTS

RESULT 1  
AAW62938  
ID AAW62938 standard; Peptide; 23 AA.

XX AC AAW62938;

XX DT 22-MAR-2001 (first entry)

XX DE Prostate homing antimicrobial pro-apoptotic conjugate.

XX KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;

XX KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;

XX KW alpha-helix; human.

XX OS Chimeric - Homo sapiens.

XX OS Chimeric - Unidentified.

XX FH Key Location/Qualifiers

XX FT Misc-difference 10..23 /note= "Preferably D-form residues"

XX PN WO200042973-A2.

XX PD 27-JUL-2000.

XX PF 21-JAN-2000; 2000WO-US01602.

XX PR 22-JAN-1999; 99US-0235902.

XX PA (BURN-) BURNHAM INST.

XX PI Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;

XX DX WPI; 2000-499174/44.

XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
 PT selectively homes to a mammalian cell type or tissue linked to an  
 PT antimicrobial peptide, useful for the treatment of prostate cancer -  
 XX  
 PS Claim 29; Page 108; 118pp; English.  
 XX  
 CC The present invention relates to homing pro-apoptotic conjugates,  
 CC comprising of a tumour homing molecule that selectively homes to a  
 CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
 CC homing pro-apoptotic conjugates are selectively internalised by the  
 CC mammalian cell type or tissue and exhibits high toxicity, especially to  
 CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
 CC toxicity when not linked to the tumor homing molecule. In addition, the  
 CC antimicrobial peptide has an amphipathic alpha-helical structure. The  
 CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,  
 CC breast and prostate cancer or melanoma. The present sequence is one such  
 CC prostate homing pro-apoptotic conjugate.  
 XX  
 SQ Sequence 23 AA;

Query Match 100.0%; Score 104; DB 21; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-09;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARLGKIAKLAKLAKLAK 23  
 Db 1 smsiarlggkiaklaklaklak 23

RESULT  
 AAE06513

ID AAE06513 standard; peptide; 23 AA.

AC AAE06513;

DT 25-SEP-2001 (first entry)

DE Chimeric prostate-homing pro-apoptotic peptide.

XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
 KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
 KW cytostatic.

XX Unidentified.

XX Key Location/Qualifiers

FT Domain 1..7

FT /label= Prostate\_homing\_domain

FT Domain 8..9

FT /label= Coupling\_domain

FT /note= "Glycinyglycine bridge"

FT Domain 10..23

FT /label= Antimicrobial\_peptide

XX WO200153342-A1.

XX 26-JUL-2001.

XX 16-JAN-2001; 2001WO-US01362.

XX 21-JAN-2000; 2000US-0489582.

XX (BURN-) BURNHAM INST.

XX Ruoslahti EI, Pasqualini R, Arap W, Bredesen DE, Ellerby HM;

XX WPI; 2001-451901/48.

XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat

PT prostate cancer, comprises a prostate-homing peptide linked to an

PT antimicrobial peptide -

XX Claim 6; Page 103; 176pp; English.  
 XX  
 CC The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used  
 CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is a chimeric  
 CC prostate-homing pro-apoptotic peptide.

XX Sequence 23 AA;

Query Match 100.0%; Score 104; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-09;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SMSIARLGKIAKLAKLAKLAK 23  
 Db 1 smsiarlggkiaklaklaklak 23

RESULT 3

AAB21936

ID AAB21936 standard; Peptide; 21 AA.

XX AAB21936;

XX 22-MAR-2001 (first entry)

XX Homing antimicrobial pro-apoptotic conjugate #1.

XX Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
 KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;  
 KW alpha-helix; human.

XX Chimeric - Homo sapiens.

XX Chimeric - Unidentified.

XX Key Location/Qualifiers

XX Misc-difference 8..21 /note= "Preferably D-form residues"

XX WO200042973-A2.

XX 27-JUL-2000.

XX 21-JAN-2000; 2000WO-US01602.

XX 22-JAN-1999; 99US-0235902.

XX (BURN-) BURNHAM INST.

XX Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;

XX WPI; 2000-499174/44.

XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
 PT selectively homes to a mammalian cell type or tissue linked to an  
 PT antimicrobial peptide, useful for the treatment of prostate cancer -

XX Claim 13; Page 105; 118pp; English.

XX The present invention relates to homing pro-apoptotic conjugates,  
 CC comprising of a tumour homing molecule that selectively homes to a  
 CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
 CC homing pro-apoptotic conjugates are selectively internalised by the  
 CC mammalian cell type or tissue and exhibits high toxicity, especially to  
 CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell



CC toxicity when not linked to the tumor homing molecule. In addition, the  
 CC antimicrobial peptide has an amphipathic alpha-helical structure. The  
 CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,  
 CC breast and prostate cancer or melanoma. The present sequence is one such  
 CC homing pro-apoptotic conjugate.

XX  
 SQ Sequence 21 AA;

Query Match 75.0%; Score 78; DB 21; Length 21;  
 Best Local Similarity 94.4%; Pred. No. 4.2e-05;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 RLGGKLAIAKLAKLAK 23  
 | | | | | | | | | | | | | | | | | | | |  
 Db 4 rcggklakiaklakiak 21

RESULT 4  
 AAE06512  
 ID AAE06512 standard; peptide; 21 AA.

XX  
 AC AAE06512;

XX  
 DT 25-SEP-2001 (first entry)

XX  
 DE Prostate-homing pro-apoptotic peptide, HPP-1.

XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;  
 KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;  
 KW cytostatic; Hpp-1.

XX  
 OS Unidentified.

XX  
 FH Key Location/Qualifiers

FT Domain 1..5

FT Disulfide-bond 1..5 Homing\_domain

FT Domain 6..7

FT /label= Coupling\_domain

FT /note= "Glycylglycine bridge"

FT Domain 8..21

FT /label= Membrane-disrupting\_domain

FT /note= "Antimicrobial peptide; This region forms an  
 amphipathic helix and is useful in imparting increased  
 stability of the conjugate in vivo"

XX  
 PN WO200153342-A1.

XX  
 PD 26-JUL-2001.

XX  
 PF 16-JAN-2001; 2001WO-US01362.

XX  
 PR 21-JAN-2000; 2000US-0489582.

XX  
 PA (BURN-) BURNHAM INST.

XX  
 PI Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;

XX  
 DR WPI; 2001-451901/48.

XX  
 PT Novel chimeric prostate-homing pro-apoptotic peptide, used to treat  
 PT prostate cancer, comprises a prostate-homing peptide linked to an  
 PT antimicrobial peptide -

XX  
 PS Example 2; Fig 1; 176pp; English.

XX The patent discloses novel chimeric prostate-homing pro-apoptotic  
 CC peptide which comprises a prostate-homing peptide linked to an  
 CC antimicrobial peptide, where the chimeric peptide is selectively  
 CC internalised by and exhibits high toxicity to prostate tissue and  
 CC where the antimicrobial peptide has low mammalian cell toxicity when  
 CC not linked to prostate-homing peptide. The chimeric peptide is used

CC to direct an antimicrobial peptide in vivo to a prostate cancer, to  
 CC induce selective toxicity in vivo in a prostate cancer, and to treat  
 CC a patient with prostate cancer. The present sequence is chimeric  
 CC prostate-homing pro-apoptotic peptide, HPP-1.

XX  
 SQ Sequence 21 AA;

Query Match 75.0%; Score 78; DB 22; Length 21;  
 Best Local Similarity 94.4%; Pred. No. 4.2e-05;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 RLGGKLAIAKLAKLAK 23  
 | | | | | | | | | | | | | | | | | | | |  
 Db 4 rcggklakiaklakiak 21

RESULT 5  
 AAB21939  
 ID AAB21939 standard; Peptide; 21 AA.

XX  
 AC AAB21939;

XX  
 DT 22-MAR-2001 (first entry)

XX  
 DE Homing antimicrobial pro-apoptotic conjugate #3.

XX Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;  
 KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;  
 KW alpha-helix; human.

XX  
 OS Chimeric - Homo sapiens.

XX  
 OS Chimeric - Unidentified.

XX  
 FH Key Location/Qualifiers

FT Misc-difference 8..21

FT /note= "Preferably D-form residues"

XX  
 PN WO200042973-A2.

XX  
 PD 27-JUL-2000.

XX  
 PF 21-JAN-2000; 2000WO-US01602.

XX  
 PR 22-JAN-1999; 99US-0235902.

XX  
 PA (BURN-) BURNHAM INST.

XX  
 PI Ellerby HM, Bredeesen DE, Pasqualini R, Ruoslahti EI;

XX  
 DR WPI; 2000-499174/44.

XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that  
 PT selectively homes to a mammalian cell type or tissue linked to an  
 PT antimicrobial peptide, useful for the treatment of prostate cancer -

XX  
 PS Disclosure; Page 8; 118pp; English.

XX The present invention relates to homing pro-apoptotic conjugates,  
 CC comprising of a tumour homing molecule that selectively homes to a  
 CC mammalian cell type or tissue, linked to an antimicrobial peptide. The  
 CC homing pro-apoptotic conjugates are selectively internalised by the  
 CC mammalian cell type or tissue and exhibits high toxicity, especially to  
 CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell  
 CC toxicity when not linked to the tumor homing molecule. In addition, the  
 CC antimicrobial peptide has an amphipathic alpha-helical structure. The  
 CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,  
 CC breast and prostate cancer or melanoma. The present sequence is one such  
 CC homing pro-apoptotic conjugate.

XX  
 SQ Sequence 21 AA;



```
AAB21940
ID AAB21940 standard; Peptide; 25 AA.
AC AAB21940;
XX
DT 22-MAR-2001 (first entry)
XX
DE Homing antimicrobial pro-apoptotic conjugate #4.
XX
KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;
KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphiathic;
KW alpha-helix; human.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT Misc-difference 12..25
FT /note= "Preferably D-form residues"
XX
PN WO200042973-A2.
XX
PD 27-JUL-2000.
XX
PF 21-JAN-2000; 2000WO-US01602.
XX
PR 22-JAN-1999; 9905-0235902.
XX
PA (BURN-) BURNHAM INST.
XX
PI Ellerby HM, Bredeesen DE, Pasqualini R, Ruoslahti EI;
XX WPI; 2000-499174/44.
XX
PT Homing pro-apoptotic conjugate comprising a tumor homing molecule that
PT selectively homes to a mammalian cell type or tissue linked to an
PT antimicrobial peptide, useful for the treatment of prostate cancer -
XX
PS Disclosure; Page 8; 118pp; English.
XX
CC The present invention relates to homing pro-apoptotic conjugates,
CC comprising of a tumour homing molecule that selectively homes to a
CC mammalian cell type or tissue, linked to an antimicrobial peptide. The
CC homing pro-apoptotic conjugates are selectively internalised by the
CC mammalian cell type or tissue and exhibits high toxicity, especially to
CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell
CC toxicity when not linked to the tumor homing molecule. In addition, the
CC antimicrobial peptide has an amphipathic alpha-helical structure. The
CC conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,
CC breast and prostate cancer or melanoma. The present sequence is one such
CC homing pro-apoptotic conjugate.
XX
SQ Sequence 25 AA;

Query Match 71.2%; Score 74; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GGKLAKLAKLAKLAK 23
DB 10 ggklaklaklaklak 25

RESULT 9
AAE06517
ID AAE06517 standard; peptide; 25 AA.
XX
AC AAE06517;
XX
DT 25-SEP-2001 (first entry)
XX
DE Homing pro-apoptotic peptide #4.
```

```
XX
KW Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;
KW antimicrobial peptide; prostate cancer; breast tumour homing molecule;
KW cytostatic.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Domain 1..9
FT /label= Homing_domain
FT Domain 10..11
FT /label= Coupling_domain
FT /note= "Glycylglycine bridge"
FT Domain 12..25
FT /label= Antimicrobial_peptide
XX
PN WO200153342-A1.
XX
PD 26-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01362.
XX
PR 21-JAN-2000; 2000US-0489582.
XX
PA (BURN-) BURNHAM INST.
XX
PI Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;
XX WPI; 2001-451901/48.
XX
PT Novel chimeric prostate-homing pro-apoptotic peptide, used to treat
PT prostate cancer, comprises a prostate-homing peptide linked to an
PT antimicrobial peptide -
XX
PS Example 3; Page 82; 176pp; English.
XX
CC The patent discloses novel chimeric prostate-homing pro-apoptotic
CC peptide which comprises a prostate-homing peptide linked to an
CC antimicrobial peptide, where the chimeric peptide is selectively
CC internalised by and exhibits high toxicity to prostate tissue and
CC where the antimicrobial peptide has low mammalian cell toxicity when
CC not linked to prostate-homing peptide. The chimeric peptide is used
CC to direct an antimicrobial peptide in vivo to a prostate cancer, to
CC induce selective toxicity in vivo in a prostate cancer, and to treat
CC a patient with prostate cancer. The present sequence is a homing pro-
CC apoptotic peptide. This peptide inhibits retinal neovascularisation.
XX
SQ Sequence 25 AA;

Query Match 71.2%; Score 74; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GGKLAKLAKLAKLAK 23
DB 10 ggklaklaklaklak 25

RESULT 10
AAB21937
ID AAB21937 standard; Peptide; 26 AA.
XX
AC AAB21937;
XX
DT 22-MAR-2001 (first entry)
XX
DE Homing antimicrobial pro-apoptotic conjugate #2.
XX
KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;
KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphiathic;
KW alpha-helix; human.
XX
```

```

OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT Misc-difference 13..26
FT /note= "Preferably D-form residues"
XX
PN WO200042973-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-US01602.
XX
XX 22-JAN-1999; 99US-0235902.
XX
XX (BURN-) BURNHAM INST.
XX
XX Ellerby HM, Bredeesen DE, Pasqualini R, Ruoslahti EI;
PI WPI; 2000-499174/44.
XX
XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that
XX selectively homes to a mammalian cell type or tissue linked to an
XX antimicrobial peptide, useful for the treatment of prostate cancer -
XX
XX Claim 13; Page 105; 118pp; English.
XX
XX The present invention relates to homing pro-apoptotic conjugates,
XX comprising of a tumour homing molecule that selectively homes to a
XX mammalian cell type or tissue, linked to an antimicrobial peptide. The
XX homing pro-apoptotic conjugates are selectively internalised by the
XX mammalian cell type or tissue and exhibits high toxicity, especially to
XX angiogenic vasculature. The antimicrobial peptide has low mammalian cell
XX toxicity when not linked to the tumor homing molecule. In addition, the
XX antimicrobial peptide has an amphipathic alpha-helical structure. The
XX conjugates are useful for the treatment of cancer e.g. Kaposi's sarcoma,
XX breast and prostate cancer or melanoma. The present sequence is one such
XX homing pro-apoptotic conjugate.
XX
XX Sequence 26 AA;
XX
XX Query Match 71.2%; Score 74; DB 21; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.0002;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 8 GGKLA KLAKLA KLAK 23
XX |||||||
XX Db 11 ggklaklaklaklak 26
XX
XX RESULT 11
XX AAE06516
XX ID AAE06516 standard; peptide; 26 AA.
XX
XX AC AAE06516;
XX
XX 25-SEP-2001 (first entry)
XX
XX Homing pro-apoptotic peptide #3.
XX
XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;
XX antimicrobial peptide; prostate cancer; breast tumour homing molecule;
XX cytostatic.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX Domain 1..10
XX FT /label= Homing_domain
XX FT 11..12
XX FT /label= Coupling_domain
XX FT /note= "Glycylglycine bridge"
XX

```

```

FT Domain 13..26
FT /label= Antimicrobial_peptide
XX
XX WO200153342-A1.
XX
XX 26-JUL-2001.
XX
XX 16-JAN-2001; 2001WO-US01362.
XX
XX 21-JAN-2000; 2000US-0489582.
XX
XX (BURN-) BURNHAM INST.
XX
XX Ruoslahti EI, Pasqualini R, Arap W, Bredeesen DE, Ellerby HM;
XX WPI; 2001-451901/48.
XX
XX Novel chimeric prostate-homing pro-apoptotic peptide, used to treat
XX prostate cancer, comprises a prostate-homing peptide linked to an
XX antimicrobial peptide -
XX
XX Example 2; Page 80; 176pp; English.
XX
XX The patent discloses novel chimeric prostate-homing pro-apoptotic
XX peptide which comprises a prostate-homing peptide linked to an
XX antimicrobial peptide, where the chimeric peptide is selectively
XX internalised by and exhibits high toxicity to prostate tissue and
XX where the antimicrobial peptide has low mammalian cell toxicity when
XX not linked to prostate-homing peptide. The chimeric peptide is used
XX to direct an antimicrobial peptide in vivo to a prostate cancer, to
XX induce selective toxicity in vivo in a prostate cancer, and to treat
XX a patient with prostate cancer. The present sequence is a homing pro-
XX apoptotic peptide.
XX
XX Sequence 26 AA;
XX
XX Query Match 71.2%; Score 74; DB 22; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.0002;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 8 GGKLA KLAKLA KLAK 23
XX |||||||
XX Db 11 ggklaklaklaklak 26
XX
XX RESULT 12
XX AAW62926
XX ID AAW62926 standard; peptide; 21 AA.
XX
XX AC AAW62926;
XX
XX 02-OCT-1998 (first entry)
XX
XX Minimalist lytic peptide.
XX
XX Lytic peptide; channel forming peptide; antibacterial; amphipathic.
XX Synthetic.
XX
XX US5789542-A.
XX
XX 04-AUG-1998.
XX
XX 06-OCT-1997; 97US-0944133.
XX
XX 22-APR-1994; 94US-0232525.
XX
XX 22-JUL-1996; 96US-0681075.
XX
XX 03-FEB-1997; 97US-0789077.
XX
XX 06-OCT-1997; 97US-0944133.
XX
XX (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.
XX

```

PI Becker CL, McLaughlin ML;  
 XX WPI; 1998-446183/38.  
 XX Selective lysis of bacteria amongst mammalian cells - using 14-mer  
 PT or 21-mer lytic peptides  
 XX  
 PS Disclosure; Column 5; 25pp; English.  
 XX  
 CC AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal  
 CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively  
 CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha -helix, whereby the multimer is amphipathic.  
 XX  
 SQ Sequence 21 AA;

Query Match 69.2%; Score 72; DB 19; Length 21;  
 Best Local Similarity 80.0%; Pred. No. 0.00032;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGKGLAKKLAKLAK 23  
 : : : | | | | | | | | | |  
 Db 2 laklakkklakklaklak 21

RESULT 13  
 AAW62943  
 ID AAW62943 standard; peptide; 28 AA.  
 XX  
 AC AAW62943;  
 XX  
 DT 02-OCT-1998 (first entry)  
 XX  
 DE Minimalist lytic peptide.  
 XX  
 KW Lytic peptide; channel forming peptide; antibacterial; amphipathic.  
 XX  
 OS Synthetic.  
 XX  
 PN US5789542-A.  
 XX  
 PD 04-AUG-1998.  
 XX  
 PF 06-OCT-1997; 97US-0944133.  
 XX  
 PR 22-APR-1994; 94US-0232525.  
 XX  
 PR 22-JUL-1996; 96US-0681075.  
 XX  
 PR 03-FEB-1997; 97US-0789077.  
 XX  
 PR 06-OCT-1997; 97US-0944133.  
 XX  
 PA (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.  
 XX  
 PI Becker CL, McLaughlin ML;  
 XX  
 WPI; 1998-446183/38.  
 XX  
 XX Selective lysis of bacteria amongst mammalian cells - using 14-mer  
 PT or 21-mer lytic peptides  
 XX  
 PS Disclosure; Column 6; 25pp; English.  
 XX  
 CC AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal  
 CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively

CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha -helix, whereby the multimer is amphipathic.  
 XX  
 SQ Sequence 28 AA;

Query Match 69.2%; Score 72; DB 19; Length 28;  
 Best Local Similarity 80.0%; Pred. No. 0.00042;  
 Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGKGLAKKLAKLAK 23  
 : : : | | | | | | | | | |  
 Db 3 laklakkklakklaklak 22

RESULT 14  
 AAW62923  
 ID AAW62923 standard; peptide; 28 AA.  
 XX  
 AC AAW62923;  
 XX  
 DT 02-OCT-1998 (first entry)  
 XX  
 DE Minimalist lytic peptide.  
 XX  
 KW Lytic peptide; channel forming peptide; antibacterial; amphipathic.  
 XX  
 OS Synthetic.  
 XX  
 PN US5789542-A.  
 XX  
 PD 04-AUG-1998.  
 XX  
 PF 06-OCT-1997; 97US-0944133.  
 XX  
 PR 22-APR-1994; 94US-0232525.  
 XX  
 PR 22-JUL-1996; 96US-0681075.  
 XX  
 PR 03-FEB-1997; 97US-0789077.  
 XX  
 PR 06-OCT-1997; 97US-0944133.  
 XX  
 PA (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.  
 XX  
 PI Becker CL, McLaughlin ML;  
 XX  
 WPI; 1998-446183/38.  
 XX  
 XX Selective lysis of bacteria amongst mammalian cells - using 14-mer  
 PT or 21-mer lytic peptides  
 XX  
 PS Disclosure; Column 5; 25pp; English.  
 XX  
 CC AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
 CC peptides have antibacterial properties in concentrations not lethal  
 CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
 CC that comprise four nonpolar amino acid residues and three positively  
 CC charged amino acid residues, or five nonpolar amino acid residues and  
 CC two positively charged amino acid residues. The nonpolar amino acid  
 CC residues and the positively charged amino acid residues are distributed  
 CC within the heptad such that when the multimer forms an alpha-helix the  
 CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
 CC and the positively charged amino acid residues will lie on the opposite  
 CC face of the alpha -helix, whereby the multimer is amphipathic.  
 XX  
 SQ Sequence 28 AA;

Query Match 69.2%; Score 72; DB 19; Length 28;  
 Best Local Similarity 80.0%; Pred. No. 0.00042;

Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGKGLAKLAKLAK 23  
:|:| | | | | | | | | |  
Db 6 IaklakkIakIakIakIak 25

RESULT 15

AAW62927  
ID AAW62927 standard; peptide; 28 AA.

XX AC AAW62927;

XX DF 02-OCT-1998 (first entry)

XX DE Minimalist lytic peptide.

XX KW Lytic peptide; channel forming peptide; antibacterial; amphipathic.

XX OS Synthetic.

XX PN US5789542-A.

XX PD 04-AUG-1998.

XX PF 06-OCT-1997; 97US-0944133.

XX PR 22-APR-1994; 94US-0232525.

XX PR 22-JUL-1996; 96US-0681075.

XX PR 03-FEB-1997; 97US-0789077.

XX PR 06-OCT-1997; 97US-0944133.

XX PA (LOU ) UNIV LOUISIANA STATE & AGRIC & MECH COLL.

XX PI Becker CL, McLaughlin ML;

XX DR WPI; 1998-446183/38.

XX PT Selective lysis of bacteria amongst mammalian cells - using 14-mer

XX PT or 21-mer lytic peptides

XX PS Disclosure; Column 5; 25pp; English.

XX CC AAW62920-67 represent minimalist lytic (channel forming) peptides. The  
XX CC peptides have antibacterial properties in concentrations not lethal  
XX CC toward mammalian cells. The peptides are heptads (or heptad multimers)  
XX CC that comprise four nonpolar amino acid residues and three positively  
XX CC charged amino acid residues, or five nonpolar amino acid residues and  
XX CC two positively charged amino acid residues. The nonpolar amino acid  
XX CC residues and the positively charged amino acid residues are distributed  
XX CC within the heptad such that when the multimer forms an alpha-helix the  
XX CC nonpolar amino acid residues will lie on one face of the alpha-helix,  
XX CC and the positively charged amino acid residues will lie on the opposite  
XX CC face of the alpha -helix, whereby the multimer is amphipathic.

XX SQ Sequence 28 AA;

Query Match 69.2%; Score 72; DB 19; Length 28;

Best Local Similarity 80.0%; Pred. No. 0.00042;

Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 IARLGKGLAKLAKLAK 23  
:|:| | | | | | | | | |  
Db 2 IaklakkIakIakIakIak 21

Search completed: May 8, 2002, 13:34:52  
Job time: 190 sec